# Plea bargaining with multiple defendants and its deterrence effect ${ }^{\text {N }}$ 

Nathan Berg ${ }^{\text {a }}$, Jeong-Yoo Kim ${ }^{\text {b,* }}$<br>${ }^{\text {a }}$ University of Otago, PO Box 56, Dunedin 9054, New Zealand<br>${ }^{\text {b }}$ Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaenumku, Seoul 130-701, Republic of Korea

## A RTICLE INFO

## Article history:

Received 10 June 2017
Received in revised form 24 January 2018
Accepted 25 April 2018
Available online 30 April 2018

## JEL classification

K41
Keywords:
Plea bargaining
Multiple defendants
Crime deterrence


#### Abstract

This article analyzes a model of plea bargaining with multiple co-defendants. We characterize equilibrium as separating or pooling, depending on the relative importance of type-I and type-II errors. Effects of plea bargaining on criminal incentives are examined in an extended model. Contrary to the widespread perception of being "soft" on crime by weakening deterrence, we show that plea bargaining unambiguously reduces crime. The benefit of improved informational efficiency more than offsets the crime-incentivizing effect of offering discounted sentences to defendants who plea bargain. Plea bargaining is therefore socially efficient whenever the risk of wrongfully convicting innocent defendants is sufficiently small.


© 2018 Elsevier Inc. All rights reserved.

## 1. Introduction

In the United States, roughly $97 \%$ of federal cases and $94 \%$ of state cases are settled by plea bargaining (Goode, 2012). In contrast, France's negotiated guilty-plea procedure introduced in 2004, is used in only $4 \%$ of decisions by the correctional courts (French Ministry of Justice, 2006). Another dimension along which plea bargaining institutions vary widely across countries is restrictions on their use. Differences across countries in frequencies of use and restrictions on plea bargaining, to a large extent, reflect conflicting prescriptive views about whether plea bargaining is socially desirable.

Although many accept the claim that plea bargaining can (at least in theory) achieve substantial improvements in informational efficiency, criticism of plea bargaining is widespread. Some argue that plea bargaining is unfair because it leads to inconsistent pun-

[^0]ishment for the same crime. Closely related is the complaint that criminals who accept plea bargain offers may not receive punishment commensurate with the crime they committed. These arguments are incomplete, however, because they consider only the effect of plea bargaining on one of two different types of judicial errors, ignoring its socially desirable effect of reducing the likelihood of convicting an innocent suspect (type-I error). We refer to social losses from under-punishment of guilty defendants as type-II error and social losses from excessive punishment of innocent defendants as type-I error. Another common criticism of plea bargaining is that it may weaken the deterrent effect of punishment by reducing expected sentences, thereby incentivizing criminal activity (Guidorizzi, 1998). Defendants who reject plea bargain offers tend to receive more severe punishments at trial than what was offered by the prosecution under plea bargaining, sometimes referred to as the "trial penalty" which, once again, attracts vehement criticism.

Notwithstanding these arguments against plea bargaining, its constitutionality was established by Brady vs. United States (1970). Since that precedent, attitudes toward plea bargaining shifted. The views of legal scholars and practitioners who, at first, regarded plea bargaining as a transient anomaly that was expected to eventually fade away later evolved into a heterogeneous majority that, despite the criticisms, accepted (perhaps begrudgingly) plea bargaining as firmly ensconced within criminal law.

Economic rationales contribute in important ways to debates over plea bargaining. Presumably, the first economic analysis of
plea bargaining is Landes (1971), arguing that plea bargaining can save time and other variable costs associated with going to trial. In Landes's model, the prosecutor's problem is specified as maximization of the number of convictions subject to a budget constraint. Landes' analysis shows that plea bargaining can improve efficiency, especially when trial convictions are costly. Informational issues and strategic thinking inherent in criminal cases are not addressed in Landes' analysis, however.

Grossman and Katz's (1983) pioneering work provides what is likely the first game-theoretic model of plea bargaining, addressing one important informational issue by providing conditions under which guilty defendants reveal their types (guilty or innocent). Grossman and Katz identify two benefits of plea bargaining-an insurance effect and a screening effect. They assert that a riskaverse defendant and a prosecutor prefer the sure conviction of a guilty defendant under plea bargaining over the risk of litigating in court. They also show that plea bargaining serves as a screening mechanism, which confers the potential advantage of improved accuracy of sentencing. Grossman and Katz characterize both a pooling equilibrium and a separating equilibrium. In the pooling equilibrium, the prosecutor makes a plea offer that is rationally accepted by both guilty and innocent defendants. In a separating equilibrium, however, the prosecutor makes an offer that is accepted by a guilty defendant but rejected by an innocent one. Thus, the typical "adverse selection" (or "lemons" problem) occurs in which defendants who accept plea offers have higher average culpability or are more likely to be guilty.

Most economic analyses of plea bargaining, however, focus on settings with only one defendant despite there being numerous examples of real-world criminal cases in which multiple co-defendants would appear to play an important role in determining sentencing outcomes. One such example is the case of price fixing and related collusive activities among multiple firms. Indeed, there are some exceptions such as Kobayashi (1992) and Kim (2009) that consider the situation of multiple co-defendants who are known to be connected with the same crime. However, neither of those papers addressed the dynamic effect of plea bargaining on crime deterrence in a multiple-defendant setting. ${ }^{1}$ This lacuna in the extant literature is rather surprising given that one of the main criticisms against plea bargaining is that it weakens crime deterrence by offering defendants shorter sentences. Reinganum (1993) and Miceli (1996) analyze the dynamical issue in a model with a single defendant, examining how plea bargaining influences the incentive to commit crimes. In this paper, we consider a dynamic model of plea bargaining between a prosecutor and multiple codefendants. In our model, defendants are not ex ante known to be guilty. The guilt or innocence of defendants is endogenously determined as the result of their criminal decisions. To investigate the effect of plea bargaining on the incentive to commit a crime, we first consider a model in which the prosecutor is unsure about whether the defendants are guilty or innocent. Because guilt is uncertain from the prosecutor's view, a socially benevolent prosecutor is assumed to pursue the objective of minimizing judicial errors, specified as a weighted sum of type-I and type-II errors (i.e., not simply maximizing penalties as in some previous models). Then, based on the analysis, we will examine the effect of plea bargaining on the incentive to commit crimes.

[^1]The objective that prosecutors in our model use-to minimize a weighted sum of losses from the two types of judicial errors-can be interpreted as consistent with guidelines codifying appropriate prosecutorial behavior across a broad range of real-world judicial systems. For example, prosecutors' duty in the US criminal justice system is to "seek justice" rather than merely convictions. ${ }^{2}$ Similarly, Article 1 of Korea's Code for Prosecutors (as instructed by the Korean Ministry of Justice) guides prosecutors to represent the public interest to minimize judicial errors-as do other countries' judicial codes-contrary to the widespread perception that prosecutors are incentivized solely to pursue the objective of maximizing penalties.

It is well known in the case of a single defendant that the prosecutor offers the defendant's certainty equivalent (i.e., the offer that gives the defendant the same disutility he expects by going to trial). In the case of multiple defendants, however, calculating multiple certainty equivalent offers is less straightforward because they depend on whether the other defendant accepts his respective offer or not. Unlike models with a single defendant, our model with multiple co-defendants allows the prosecutor to make plea offers contingent on the defendant's promise to both plead guilty and testify against the other co-defendant. Therefore, a certainty equivalent offer to one defendant must be contingent on whether the other defendant accepts his offer or not. We characterize all possible separating-equilibrium offers that are accepted only by guilty defendants and all pooling offers that are accepted by both guilty and innocent defendants. ${ }^{3}$

In both types of equilibria, plea offers must be fair in the sense that the more culpable defendant (i.e., the one who deserves a longer sentence) receives a harsher penalty, unless both defendants are equally culpable. Our model's result that any plea bargaining equilibrium with multiple co-defendants must necessarily respect at least this rather weak notion of fairness stands in sharp contrast to Kobayashi's (1992) model in which unfair equilibria are possible (i.e. in which the more culpable defendant may receive a less severe penalty). ${ }^{4}$ These contrasting predictions regarding the fairness of equilibrium plea bargaining in our model and Kobayashi's (1992) are the result of different equilibrium concepts. Unfair equilibria are impossible in our model as long as the spirit of Nash equilibrium is respected by implicitly requiring all agents' beliefs to be consistent with their equilibrium strategies.

In a separating equilibrium, the plea bargain offers are asymmetric in that only the less culpable defendants are offered a plea discount. In a pooling equilibrium, both defendants are offered plea discounts. Intuitively, longer (i.e., more severe) plea offers in the pooling equilibrium, as they approach the duration for a guilty defendant without plea bargaining, increase the loss from typeI errors when a defendant is actually innocent (excessively harsh sentencing for the innocent defendants) and decrease the loss from type-II errors (insufficiently harsh sentencing for the guilty defendants). The optimal pooling equilibrium in our model is determined by the plea offer that balances these two effects (i.e., equating the marginal benefit of reducing type-II error with the marginal cost of increasing type-I error for each defendant). The prosecutor's choice between the separating or pooling equilibrium depends on the relative importance of type-I and type-II errors. If type-I errors are sufficiently more important in the prosecutor's objective function,

[^2]then optimal plea offers are characterizable as those that convey larger aggregate rewards offered to, and accepted by, both types of defendants, which leads to the pooling equilibrium. In the other direction, the separating equilibrium is selected whenever type-II errors receive greater relative weight in the prosecutor's objective function.

The result that only one defendant receives a plea discount in a separating equilibrium holds even when two codefendants are equally culpable. If the prosecutor gives plea discounts to both defendants, type-II errors become much higher than when he gives a plea discount to only one defendant, making the other accept an undiscounted offer. The former case (allowing discounts to both defendants) is equivalent to the case that the prosecutor makes plea bargaining with only one defendant twice. This shows an important difference between plea bargaining with a single defendant and with multiple defendants.

It is worthwhile further comparing our result with that of Kobayashi (1992). Kobayashi considers a complete-information game of plea bargaining with two co-defendants. In his model, it is known that both defendants are guilty. Therefore, the objective of Kobayashi's prosecutor is to maximize the sum of the two defendants' expected penalties (or, equivalently, minimize the sum of expected discounts, i.e., the aggregate reward, relative to the expected punishment without plea bargaining). Kobayashi obtains the counterintuitive result that a more culpable defendant may receive a more lenient penalty, which is intuitively unfair. Such unfair outcomes occur in his model when, for example, a more culpable Defendant 1 believes that the less culpable Defendant 2 is very unlikely to accept the offer made to Defendant 2 , while Defendant 2 believes that Defendant 1 is highly likely to accept his respective offer. In this case, because of the role of the other defendant's exogenously given belief in the best-response function, Defendant 1 may rationally choose to be more likely to reject the offer than Defendant 2. Thus, the prosecutor is forced to make a more attractive offer to the more culpable Defendant 1 to induce him to accept his respective offer, implying that the offer to Defendant 1 can-in theory-be lower than the offer to less culpable Defendant 2 in equilibrium. This unsettling possibility of the less culpable defendant receiving a more severe punishment follows from Kobayashi's assumption that each defendant's belief regarding the other's acceptance decision is exogenously given. However, once we require that all players' beliefs satisfy the condition of being consistent with defendants' equilibrium strategies, then Kobayashi's unfair plea bargains cannot occur in any equilibrium in our model.

Another difference between Kobayashi's model and ours is the meaning of "more culpable." We define a defendant as being more culpable if he deserves a longer sentence than the other. In contrast, Kobayashi defines a defendant as more culpable if he has a higher probability of conviction and his acceptance of the plea offer increases the probability of conviction by more than acceptance of the plea offer by the other defendant does. In our model, acceptance of the plea offer by the more culpable defendant does not necessarily imply a larger increase in the probability that the other defendant is convicted. Both defendants' conviction probabilities are increased equally no matter which co-defendant accepts the plea offer. In other words, both defendants have the same information regardless of which co-defendant is more culpable. Because the more culpable defendant's sentence conditional on going to court is expected to be greater, this more culpable defendant benefits more from judicial errors. Consequently, the prosecutor must offer larger plea-bargaining discounts to a more culpable defendant (who is less willing to accept plea offers), in order to induce him to accept it. Since larger discounts (i.e., reductions in sentencing) are costly from the prosecutor's point of view, the prosecutor chooses
to offer a plea discount only to the less culpable defendant. Hence, equilibrium offers must be fair.

Despite the advantage of static efficiency achieved by reducing judicial errors through plea bargaining, there remain important concerns about the possibility of dynamic inefficiency insofar as reduced sentencing may increase the incentive to commit crimes. One of the strongest criticisms against plea bargaining, which by many accounts would appear to rest upon a theoretically obvious mechanism, avers that plea bargaining can lead to softer sentencing, thereby weakening the deterrent effect of expected punishment and consequently increasing crime rates. Although this mechanism may seem obvious (i.e., that potential criminals would, as defendants, face lower expected sentences under plea bargaining and therefore have a greater likelihood of committing a crime), the plea bargaining mechanism does indeed have a crime-deterrent effect. In a separating equilibrium, the benefit of increased information revealed by the plea bargaining institution more than offsets the cost of weakened deterrence (e.g., reductions in the mean duration of sentences), which (in any separating equilibrium) we show is degenerate at zero; thus, the plea bargaining institution, in our model's separating equilibrium, leads unambiguously to lower crime rates. Intuitively, this is because plea bargaining allows the prosecutor to collect incriminating evidence against co-defendants based on their testimonies. In a pooling equilibrium as well, plea bargaining provides stronger deterrence than without plea bargaining. A similar intuition can be applied to a pooling equilibrium. The prosecutor can make harsher plea offers acceptable by using informational enhancement as a threat strategy.

The deterrent effect of plea bargaining was analyzed earlier by Reinganum (1993) and Miceli (1996) but only for a single defendant. Miceli's model is similar to ours in that it focuses on a screening mechanism. Miceli (1996) considers a two-stage game in which a legislature first determines criminal punishments and then, once crimes have been committed as a best response to the statutory punishments, actual punishments are determined by prosecutors and judges in the subsequent stage of the game with plea-bargaining or going to trial. Miceli shows that if the legislature raises the magnitude of punishment too high, then the deterrent effect may counterintuitively be weakened due to the response of prosecutors who believe that severe statutory punishments do not fit the crime. Reinganum (1993) considers a two-stage signaling game based on Reinganum (1988). Our model extends Miceli's analysis to the case of two co-defendants although our model has several important differences.

Our finding that plea bargaining with multiple co-defendants unambiguously deters crime sheds new light on debates over the informational and social efficiency of real-world leniency programs. One of the important policy questions in recent decades concerns how to revise leniency programs that grant amnesty (e.g., to firms that previously engaged in illegal antitrust activity) to promote improvements in social efficiency. Because many illegal antitrust activities such as price-fixing involve more than a single firm, our model of plea bargaining with multiple co-defendants can provide novel insight relevant to the design of optimal leniency programs.

The article is organized as follows. In Section 2, we specify the basic setup for our model of plea bargaining with multiple codefendants under incomplete information. In Section 3, taking the plea offers as given, we analyze the defendants' decisions about whether to accept pleas offers. In Section 4, we characterize the separating equilibrium offers and pooling equilibrium offers by the prosecutor. In Section 5, we discuss some modifications of our plea bargaining model. In Section 6, we consider an extension that brings in defendants' decisions about whether to commit a crime by adding an earlier stage of the game before plea bargaining begins.

Section 7 contains concluding remarks and further caveats. Proofs are relegated to the Appendix.

## 2. Model

We closely follow the model of Kim (2009). Thus, the model we present in this section is basically an incomplete information version of the model in Kim (2009).

Suppose that prosecutor $P$ has accused two co-defendants, $D_{i}$, $i=1,2$, of jointly committing a crime. The co-defendants are either guilty ( $G$ ) or innocent ( $I$ ). Defendants are referred to as type $t=G$ or $I$, denoted $D_{i}(G)$ or $D_{i}(I)$ respectively, if they are guilty or innocent. Each defendant's type $t$ is known by both defendants but not by the prosecutor. $P$ is assumed to know only that $\operatorname{Prob}(t=G)=\gamma \in(0$, $1)$. The prosecutor's belief probability $\gamma$ is assumed to be common knowledge among all players including the judge ( $J$ ).

Sentences $s_{i}(>0)$ are ordered by the judge for each defendant convicted of committing the crime. Without loss of generality, the two sentences are assumed to satisfy the inequality $s_{1} \geq s_{2}$, which describes a situation where one defendant receives a weakly larger sentence than the other. ${ }^{5}$ We will say that $D_{1}$ is more culpable than $D_{2}$ if they are guilty and $s_{1}>s_{2} \cdot{ }^{6}$ It is also assumed that $s_{1}$ and $s_{2}$ are common knowledge.

In the event of a plea bargain agreement, defendants may instead serve reduced sentences by accepting the prosecutor's plea offer. Usually a plea bargain offered to a defendant is made in exchange for his promise to testify in court to support a particular fact or piece of evidence. Our model focuses on plea bargaining in which the defendant agrees to plead guilty and testify against the other defendant in exchange for a reduced sentence.

The game between $P, D_{1}$ and $D_{2}$ is modeled as follows. $P$ makes simultaneous plea offers $b_{i} \in[0, \infty)$ to each defendant $D_{i}$. Each $D_{i}$ then decides whether to accept $\left(d_{i}=A\right)$ or reject the offer $\left(d_{i}=R\right)$. If $D_{i}$ accepts the offer, then his sentence in court is $b_{i}$ with certainty. If $D_{i}$ rejects the offer, then $J$ decides whether to convict him. ${ }^{7}$ In this case, $J$ may find that $t=G, I$ based on the evidence submitted at trial by $P, D_{1}$ and $D_{2}$. The probability that a defendant is convicted depends on $t$ and whether the other defendant has agreed to testify against him. Denote the conviction probability when neither accepts the offer as $q$. We assume: (i) that $q(t) \in(0,1)$ for both types, $t$; and (ii) that $q(I)<q(G)$. Assumption (i) implies that there is a strictly positive probability of both type-I and type-II errors: $q(I)>0$ implies that there are type-I errors and $q(G)<1$ implies that there are type-II errors. The assumption that the probability distributions for both types of judicial errors are non-degenerate is crucial in driving our subsequent analysis. Assumption (ii) reflects the intuition (and empirical reality) that innocent defendants can defend themselves better than guilty defendants can. ${ }^{8}$ The timeline is shown in Fig. 1.

[^3]

Fig. 1. The sequence of events.

If $D_{i}$ accepts $b_{i}$ and testifies against $D_{j}$, then we assume that $D_{j}$ is convicted with probability one. ${ }^{9}$ Indeed, $J$ might conceivably interpret a defendant's rejection of a plea bargain offer as a signal of that defendant's innocence, which we rule out, however, with the assumption that such a signal would not affect $q(t) .{ }^{10}$ Moreover, we assume that $J$ has no judicial discretion, implying that once the defendants are found to be guilty at trial, then all $J$ can do is simply order the sentence $s_{i}$ for any defendant who did not plead guilty. ${ }^{11}$ In order to isolate the informational motive for plea bargaining, which is the focus of our model, from cost-saving motives analyzed by others, our model also abstracts from trial costs by assuming them to equal zero. ${ }^{12}$

Let $\tilde{x}_{i}$ denote the correct sentence for $D_{i}$ and $x_{i}$ denote the sentence actually offered. We assume that $D_{i}$ minimizes his own expected sentence $x_{i}$. P, based on the benevolent social-welfare objective of minimizing judicial error (i.e., matching sentences to appropriately fit the crimes committed), is assumed to minimize a weighted average of losses due to type-I errors (which occur when $x_{i}>\tilde{x}_{i}$ ) and losses due to type-II errors (which occur when $x_{i}<\tilde{x}_{i} .{ }^{13}$ Thus, the defendant $D_{i}$ 's loss-function objective is:
$W_{i}\left(x_{i} ; \tilde{x}_{i}\right)=x_{i}$.
We could interpret $x_{i}$ as the loss or disutility from the sentence $x_{i}$. $P$ 's loss-function objective, denoted $L_{P}$, measures the loss associated with an accusation against two defendants as the sum of losses $L_{i}$ associated with each defendant (based on the weighted sum of type-1 and type-2 errors for each):

$$
\begin{aligned}
L_{P}\left(x_{1}, x_{2} ; \tilde{x}_{i}\right) & =\sum_{i=1}^{2} L_{i}\left(x_{i}, \tilde{x}_{i}\right), \\
L_{i}\left(x_{i} ; \tilde{x}_{i}\right) & = \begin{cases}\theta\left\|x_{i}, \tilde{x}_{i}\right\| & \text { if } x_{i}>\tilde{x}_{i} \\
(1-\theta)\left\|x_{i}, \tilde{x}_{i}\right\| & \text { if } x_{i}<\tilde{x}_{i} \\
0 & \text { if } x_{i}=\tilde{x}_{i}\end{cases}
\end{aligned}
$$

where $\|x, y\|$ is a metric measuring the distance between $x$ and $y,{ }^{14}$ and $\theta \in(0,1)$ represents the relative importance of type-I errors. For simplicity, we assume that $\|x, y\|=(x-y)^{2}$. And because the role of $J$ in our model is simply a machine that orders convictions and

[^4]Table 1
Decisions of Co-defendants.

|  |  | $D_{2}$ |  |
| :--- | :--- | :--- | :--- |
|  |  | $A$ | $R$ |
| $D_{1}$ | $A$ | $\left(b_{1}, b_{2}\right)$ | $\left(b_{1}, s_{2}\right)$ |
|  | $R$ | $\left(s_{1}, b_{2}\right)$ | $\left(q(t) s_{1}, q(t) s_{2}\right)$ |

sentences by applying rules (as specified in our model), there is no need to introduce a loss function for $J$.

We are implicitly assuming that plea-offer negotiations take place jointly, whereby both defendants can observe all plea offers, in this case both $b_{1}$ and $b_{2}$, and each defendant makes his individual decision of whether to accept or reject without having first observed the other defendant's decision. ${ }^{15}$ We can then express $D_{i}$ 's strategy as the rule $d_{i}: B \times T \rightarrow\{A, R\}$, where $B=[0, \infty)^{2}$ is the set of all possible plea offers and $T=\{G, I\}$ is the set of types.

## 3. Defendants' decision rules

As is typical of sequential games, we use backward induction. In this section, we develop the second-stage Nash-equilibrium decision rules of the two defendants conditional on the plea offers ( $b_{1}$, $b_{2}$ ) decided earlier by the prosecutor in the first stage of the game. Defendants' losses of the second-stage game are summarized in Table 1.

For each type $t$ (of both defendants, each of whom may reach different decisions about accepting plea offers), there are four possible profiles of "accept" and "reject" that code the individual decisions made by each of the two defendants. In turn, the two-dimensional space of $P$ 's plea offers can be partitioned into regions that correspond to these four possible profiles representing the defendants' joint decisions (once again, denoting these respective decisions as $A$ for "accept" and $R$ for "reject") ${ }^{16}$ :
$A A(t)=\left\{\mathbf{b}=\left(b_{1}, b_{2}\right) \in B \mid b_{i} \leq s_{i}, i=1,2\right\}$,
$R R(t)=\left\{\mathbf{b} \in B \mid b_{i}>q(t) s_{i}, i=1,2\right\}$,
$A R(t)=\left\{\mathbf{b} \in B \mid b_{1} \leq q(t) s_{1}, b_{2}>s_{2}\right\}$,
$R A(t)=\left\{\mathbf{b} \in B \mid, b_{1}>s_{1}, b_{2} \leq q(t) s_{2}\right\}$.
$A A(t)$ represents the set of plea offers that are accepted by both defendants. $R R(t)$ represents the set of plea offers that are rejected by both defendants. $A R(t)$ and $R A(t)$ are the sets of plea offers that are accepted by only one defendant. These four regions are illustrated in Fig. 2.

As shown in Fig. 2, the regions of joint acceptance and joint rejection, $A A(t)$ and $R R(t)$, overlap in some set $M=\left\{\mathbf{b} \mid q(t) s_{i}<b_{i} \leq s_{i}\right\}$. The implication of this overlap is that there are multiple equilibria: one in which all $\mathbf{b} \in M$ are accepted by both defendants; and the other in which all $\mathbf{b} \in M$ are rejected by both defendants. If each of the defendants believes that the other defendant will accept the offer, then the $A A$-equilibrium will be realized. If both defendants believe that the offer to the other will be rejected, then the $R R$-equilibrium will be realized.

The selection criterion of Pareto dominance can be used to deal with this multiplicity of equilibria. The $R R$-equilibrium Pareto dominates the $A A$-equilibrium as long as $b_{i}>q(t) s_{i}$. We apply this equilibrium selection criterion as needed. ${ }^{17}$

[^5]

Fig. 2. Mapping from prosecutor's offer space ( $b_{1}, b_{2}$ ) into two defendants' "Accept" (A) or "Reject" (R) decision profiles.

If $b_{i}<q(t) s_{i}$ for all $i=1,2$, then "accept" $\left(d_{i}=A\right)$ is the dominant strategy for defendant $i$ in the game between two defendants. However, the game differs from Prisoners' Dilemma in the sense that the non-cooperative outcome obtained when both defendants "defect" by accepting their respective plea offer is Pareto superior to the cooperative outcome obtained when both defendants "cooperate" by rejecting the offer. Hence, the strategic interaction between the two defendants is not a "dilemma." The defect-defect profile (i.e., both players accepting plea offers) is, in our model, collectively rational as well as individually rational. ${ }^{18}$

## 4. Prosecutor's decision

In Section 2, we assumed that a prosecutor minimizes a weighted sum of type-I and type-II errors. If she makes plea offers $\mathbf{b}=\left(b_{1}, b_{2}\right)$ to co-defendants, it determines her expected losses (weighted sum of errors), $L_{P}(\mathbf{b})=L_{1}(\mathbf{b})+L_{2}(\mathbf{b})$, according to the decisions of the defendants as described in Section 3.

Then, the prosecutor's equilibrium plea offer must be $\mathbf{b}^{*} \in$ $\operatorname{argmin}_{\mathbf{b} \in B} L_{P}(\mathbf{b})$. Depending on the choice of $\mathbf{b}^{*}$, defendants' types may be revealed by their decisions or remain un-revealed. Let $B^{S}$ and $B^{P}$ be the sets of plea offers that induce the type of defendants separated or pooled, respectively. Fig. 3 depicts $B^{S}$ and $B^{P}$. A separating equilibrium occurs if $\mathbf{b}^{*} \in B^{S}$ and a pooling equilibrium occurs if $\mathbf{b}^{*} \in B^{P}$. Also, we say that the plea offers $b_{1}$ and $b_{2}$ are fair whenever the condition, $b_{1}>b_{2}$ if and only if $s_{1}>s_{2}$, holds.

[^6]

Fig. 3. Separating and pooling offers given defendants' type, guilty (G) or innocent (I).


Fig. 4. Separating equilibrium.

### 4.1. Separating equilibrium

In this section, we characterize the separating equilibrium. A separating equilibrium offer of offers must lie somewhere in $B^{S}$. Note that all points in $B^{S}$ generate the same type-I error, because they are always rejected by innocent defendants, which means that $P$ 's loss is independent of the plea offers in this region. Thus, the $P$ 's most preferred point in this region, i.e., minimizing $L_{P}(\mathbf{b})$, is the one that minimizes type-II errors. The type-II errors associated with a point such as point $A$ in Fig. 4 can be depicted by an iso-loss circle that consists of offer pairs that produce the same type-II error. It is easy to see from Fig. 4 that the type-II errors decrease monotonically in $b_{1}$ until $b_{1}$ reaches $s_{1}$, and thus point $E_{1}$ yields the least type-II error among the points in $B^{S} .{ }^{19}$

Proposition 1. If the pair of plea offers $\left(b_{1}^{s}, b_{2}^{s}\right)$ is a separating equilibrium, then $\left(b_{1}^{s}, b_{2}^{s}\right)=\left(s_{1}, q(G) s_{2}\right)$.

This proposition has an important implication for real-world corporate leniency programs. The result of the proposition that optimal plea offers should reduce the sentence only to one

[^7]defendant ${ }^{20}$ supports the current US Corporate Leniency Program ${ }^{21}$ granting amnesty only to the first firm that comes forward (although there is a second aspect of that program that grants full amnesty to the first self-reporter, which is not supported by our theory in the sense that it leads to a suboptimally large judicial error).

This proposition also implies that $b_{1}^{s}>b_{2}^{s}$ if $s_{1}>s_{2}$, which suggests that the separating equilibrium offers are fair as long as $s_{1}>s_{2} .22$ One interesting feature of the equilibrium pair of offers is that those offers must be asymmetric even if $s_{1}=s_{2}$. This observation suggests that separating offers are unfair if and only if $s_{1}=s_{2}$. The intuition for this result is as follows. Only guilty defendants choose to accept the offers in a separating equilibrium. Therefore, it is better from the prosecutor's point of view for those offers to be as close as possible to the penalties that fit the crime, $(s, s)$, where $s=s_{1}=s_{2}$. Symmetric offers that are both accepted must be at most $q(G) s$. Therefore, such a pair of offers far away from $(s, s)$ can never be optimal according to $P$ 's objective, because $P$ could, in that case, always reduce the risk of type-II error by slightly increasing one of the plea offers above the other.

Why is a less culpable defendant (assuming both defendants are guilty) offered a less harsh sentence whenever $s_{1} \neq s_{2}$ ? For separation, at least one guilty defendant must accept his respective plea offer. From P's point of view, it is more costly to induce the more culpable defendant to accept the offer, because a larger plea discount is required. To see this, note that $q(G) s_{i}$ is the maximal plea offered to $D_{i}(G)$ given that $D_{j}$ rejects. Then, for the two required plea discounts, we have the inequality $s_{1}-q(G) s_{1}>s_{2}-q(G) s_{2}$, because $s_{1}>s_{2}$. This implies that $P$ always prefers to offer a more lenient plea to the less culpable of the two defendants so that it will be accepted at minimum cost. In short, it is more costly (i.e., requires larger discounts) to make a more culpable defendant accept a plea offer.

### 4.2. Pooling equilibrium

In this section, we consider the possibility of a pooling equilibrium in which defendants choose the same action regardless of their types (guilty or innocent). Although Fig. 3 indicates that $B^{P}$ consists of two disjoint regions, i.e., $B^{P}=B_{1}^{P} \cup B_{2}^{P}$ where $B_{1}^{P}=R R(G)$ (the upper northeast region of Fig. 3) and $B_{2}^{P}=A A(I) \cup A R(I) \cup R A(I)$. We can show that pooling equilibrium offers cannot be made in the region labeled $B_{1}^{P}$.

Lemma 1. None of the plea offers in $B^{P}(=R R(G))$ can be a pooling equilibrium.

The intuitive reason why pooling offers in $R R(G)$ are inferior to separating offers is as follows. Innocent types reject in either pooling or separating equilibria. Guilty types, however, accept separating plea offers and reject the pooling offers. Guilty defendants (proceeding to trial after rejecting the pooling offers) incur the cost of type-II error due to judicial error, $q(G)$. But the separating plea offer $b_{1}=s_{1}$ (without allowing plea discounts) eliminates typeII error involving the more culpable defendant, which leads $P$ to strictly prefer a separating equilibrium.

Next we consider a pair of pooling offers in the interior of $B_{2}^{P}$ such as point $A$ in Fig. 5. Such offers are accepted by both guilty

[^8]

Fig. 5. Pooling equilibrium.
and innocent defendants. Thus, type-I errors occur if $t=I$ and typeII errors occur if $t=G$. The iso-loss circles associated with type-I and type-II errors, respectively, are illustrated by circles in Fig. 5. Circle $C_{1}$ is the set of plea-offer pairs ( $b_{1}, b_{2}$ ) that yield the same level of type-I errors, and $C_{2}$ is the set of plea-offer pairs ( $b_{1}, b_{2}$ ) yielding the same level of type-II errors. It is straightforward to see from Fig. 5 that both type-I and type-II errors can be reduced by moving in the direction of the arrow from point $A$ southeast. Thus, it directly follows that point $A$ cannot be a pooling equilibrium pair of offers. Only points where the two iso-loss circles are tangent along the line $b_{2}=\frac{s_{2}}{s_{1}} b_{1}$ can be pooling offers.

Lemma 2. None of the plea offers in the interior of $A A(I)$ except offers ( $b_{1}, b_{2}$ ) satisfying $b_{2}=\frac{s_{2}}{s_{1}} b_{1}$ can be a pooling equilibrium.

Also, any point on the vertical boundary of $A A(I)$ in $B^{P}$ cannot be a pooling equilibrium pair of offers. So, we have

Lemma 3. None of the plea offers with $b_{1}=q(I) s_{1}$ and $b_{2}>q(I) s_{2}$ can be a pooling equilibrium.

The asymmetry (between the vertical and horizontal boundaries) follows from the defendants' different degrees of culpability (i.e., because $s_{1}>s_{2}$ ) and can be explained intuitively as follows. A pair of plea offers ( $b_{1}, b_{2}$ ) on the vertical boundary of $R R(I)$ would mean that $D_{1}$ is offered a relatively more advantageous discount in the sense that $\frac{s_{2}-b_{2}}{s_{1}-b_{1}}<\frac{s_{2}}{s_{1}}$. For any such point, there exists a pair of offers that is symmetric around the diagonal line given by $\frac{b_{2}}{b_{1}}=\frac{s_{2}}{s_{1}}$. These offers would give relative advantage to $D_{2}$ in the same way as before (i.e., giving $D_{1}$ a discount of $s_{2}-b_{2}$ and $D_{2}$ a discount of $s_{1}-b_{1}$ ). $P$ is indifferent between the two pairs of offers just described, one relatively more generous to $D_{1}$ and the other relatively more generous to $D_{2}$. Even if the latter pair is not feasible (in the sense that $s_{i}<0$ for some $i$ ), there must exist a point between them that can be accepted by both innocent defendants and which is strictly preferred by $P$. In other words, $P$ can symmetrically duplicate any pair of offers favorable to a more culpable defendant. Or $P$ may be able to achieve a strictly greater payoff by giving (almost) the same plea discount to a less culpable defendant, once again, because it is more costly to induce a more culpable defendant to accept.

These lemmas lead us to conclude that an equilibrium pair of offers must be located either at a tangency of two iso-loss circles associated with type-I and type-II errors (shown by the diagonal segment in Fig. 5, proven subsequently), or along the horizontal boundary of $A A(I)$.

Proposition 2. If $\left(b_{1}^{p}, b_{2}^{p}\right)$ is a pooling equilibrium pair of offers, then $\left(b_{1}^{p}, b_{2}^{p}\right)=\left(\Gamma s_{1}, \Gamma^{\prime} s_{2}\right)$ where $\Gamma \equiv \frac{\gamma(1-\theta)}{(1-\gamma) \theta+\gamma(1-\theta)}$ and $\Gamma^{\prime}=\min \{\Gamma, q(I)\}$.

The proposition above suggests that pooling equilibrium offers are always fair. Unlike separating equilibrium offers, pooling equilibrium offers can be symmetric when $s_{1}=s_{2}$.

A necessary condition for a pair of plea offers to be pooling equilibrium offers is to
$\min _{\mathbf{b} \in B^{P}} L_{P}(\mathbf{b})=L_{1}(\mathbf{b})+L_{2}(\mathbf{b})$,
where $L_{i}(\mathbf{b})=\gamma(1-\theta)\left(s_{i}-b_{i}\right)^{2}+(1-\gamma) \theta b_{i}^{2} .{ }^{23}$ Since $L_{P}(\mathbf{b})$ is additively separable, an interior solution must satisfy the two first-order conditions ( $i=1,2$ ):
$\frac{\partial L_{P}}{\partial b_{i}}=\frac{\partial L_{i}}{\partial b_{i}}=-2 \gamma(1-\theta)\left(s_{i}-b_{i}^{p}\right)+2(1-\gamma) \theta b_{i}^{p}=0$.
The first term in the left-hand-side expression above represents the marginal benefit of reduced type-II errors associated with a small increase in $b_{i}$. The second term measures the marginal cost of an increased probability of type-I error. The optimal plea offers are defined by the requirement of balancing these two effects.

From Eq. (1), one obtains the pooling equilibrium offers:
$b_{i}^{p}=\Gamma s_{i}$,
where $\Gamma \equiv \frac{\gamma(1-\theta)}{(1-\gamma) \theta+\gamma(1-\theta)} \in(0,1)$. An interior solution is possible if $\Gamma<q(I)$. Note that the set of interior solutions (satisfying the tangency condition) corresponds to the diagonal line segment defined by $\frac{b_{2}}{b_{1}}=\frac{s_{2}}{s_{1}}$.

If $\gamma \geq q(I)$, then we have corner solutions. For corner solutions (offering a culpable defendant precisely his expected sentence without plea bargaining, i.e., zero discount), it is convenient to transform the condition that $\mathbf{b} \in B^{P}$ into the condition that $b_{1} \leq s_{1}$ and $b_{2}=q(I) s_{2}$, due to Lemma 2. One of the first-order conditions given by (1) is still valid. Therefore, such a corner solution must satisfy the following conditions:
$b_{1}^{p}=\Gamma s_{1} \quad$ and $\quad b_{2}^{p}=q(I) s_{2}$.
The magnitude of $\theta$ and $\gamma$ determine whether a pooling equilibrium is an interior solution or a corner solution. If $\theta$ is very large, then type-I errors are important, so plea offers will be very low. As $\theta$ becomes smaller, or as $\gamma$ becomes larger, type-I errors become less important and thus plea offers increase along the diagonal in Fig. 5. If $\theta$ becomes even larger, then pooling offers can no longer be increased along the diagonal without inducing a separating equilibrium. For a pair of offers to remain in the pooling equilibrium region (i.e., to be both accepted by the innocent types), therefore, either $b_{1}$ or $b_{2}$ must be discounted. As long as $s_{1}>s_{2}$, however, it is less costly to discount $b_{1}$. This is exactly the intuition behind Lemma 2.

Comparing the formulas presented so far for separating and pooling equilibria, two observations are worth emphasizing. First, the pooling equilibrium offers depend on (and vary with respect to) the prosecutor's views of the relative importance of type-I errors, $\theta$, and belief about the unconditional probability of guilt, $\gamma$. In contrast, the separating equilibrium pair of plea offers presented in the previous section is, if it exists, unique (i.e., constant with respect to $\theta$ and $\gamma$ ).

It remains to see which of the separating versus pooling offers $P$ will choose. Intuitively, if $P$ believes type-I errors are more important, then she will choose low offers that can be accepted by both defendants, which leads to a pooling equilibrium. If instead

[^9]

Fig. 6. Separating and pooling equilibrium offers.
the prosecutor believes that type-II errors are more important, then she will prefer high (i.e., non-lenient, long-duration) offers so that innocent defendants reject them, which leads to the separating equilibrium. Fig. 6 illustrates the equilibrium plea offers with changes in $\Theta$ where $\Theta=\frac{(1-\gamma) \theta}{\gamma(1-\theta)}=\frac{1}{T}-1$. The figure shows that $P$ prefers a separating equilibrium if $\Theta \leq \bar{\Theta}$ (or $\Gamma \geq \underline{\Gamma}$ ), or equivalently, $\gamma \geq \frac{\theta}{\theta+(1-\theta) \bar{\Theta}} \equiv \underline{\gamma}$, where $\underline{\gamma}$ denotes the minimum value of $\gamma$ for which a separating equilibrium is supported; and prefers a pooling equilibrium otherwise.

As a concrete numerical example, we consider the parameterization for which $s_{1}=2, s_{2}=1, q(G)=.8$ and $q(I)=.4$. With exogenous parameters set to these values, $P$ prefers the separating offers over the pooling offers if $\Theta<\bar{\Theta} \approx 1.466$ (or $\Gamma>\underline{\Gamma} \approx 0.406$ ) and prefers the pooling offers otherwise.

## 5. Discussion

In this section, we discuss the effects of alternative assumptions that capture institutional details to strengthen the implications of the model.

### 5.1. The possibility of dismissing the case

It is well recognized that there is an inherent problem of time inconsistency in most of the screening models of plea bargaining originating from Grossman and Katz's seminal work. The screening outcome critically relies on the assumption that the prosecutor can credibly commit to go to trial whenever the plea offer is rejected. Contrary to this assumption, if the prosecutor infers that a defendant who rejected an offer must be innocent as predicted in our separating equilibrium, then the prosecutor would have no reason to take the case to trial. The commitment to go to trial is therefore hardly credible.

If it is known that the prosecutor will drop the case after a plea offer is rejected, then a guilty defendant would also choose to reject the offer. If the commitment to stand trial is not credible, then there cannot be a separating equilibrium that self-selects the guilty and the innocent. This disturbing outcome is shared by most of the relevant literature on plea bargaining. ${ }^{24}$ This problem can be resolved by considering mixed strategies (Baker and Mezzetti, 2001; Kim, 2010; Daughety and Reinganum, 2016).

[^10]Consider a modified game between $P, D_{1}$ and $D_{2}$ specified as follows. $P$ offers $b_{i}$ to $D_{i}, i=1,2$ and then each $D_{i}$ accepts his respective offer with probability $\alpha_{i}$ and rejects it with probability $1-\alpha_{i}$. If an offer is rejected by $D_{i}$, then $P$ decides whether to proceed to trial against $D_{i}$ with probability $\beta_{i}$ or to drop the case with probability $1-\beta_{i}$. Using such a setup, one can find a semi-separating equilibrium in which plea bargaining provides some additional information regarding the co-defendants' guilt. In this equilibrium, the prosecutor makes plea offers that are accepted only by guilty defendants with some positive probability. This result leads the prosecutor to adjust her belief about the co-defendants' guilt, thus triggering a trial with some positive probability.

Formally, we use backward induction. Consider first P's decision to dismiss the case given that the offer $b_{i}$ has been rejected. If only defendant $i$ rejects, then $P$ will go to trial against $D_{i}$ so long as it reduces expected loss (i.e., so long as the following inequality is satisfied $): E\left[L_{P} \mid\right.$ gototrial, $\left.\hat{\gamma}\right]-E\left[L_{P} \mid\right.$ dismiss, $\left.\hat{\gamma}\right]=(1-\hat{\gamma}) \theta s_{i}-\hat{\gamma}(1-$ $\theta) s_{i}=[(1-\hat{\gamma}) \theta-\hat{\gamma}(1-\theta)] s_{i}=(\theta-\hat{\gamma}) s_{i}<0$, which holds if and only if $\hat{\gamma}>\theta$. Alternatively, $P$ will be indifferent between going to trial and dropping the case if:
$\hat{\gamma}=\theta$.
We will restrict our attention to the case that $P$ chooses a completely mixed strategy, i.e., $\hat{\gamma}=\theta$, since our interest is in the existence of a semi-separating equilibrium.

On the other hand, if both offers are rejected, $P$ will compare the following two losses:
$E\left[L_{P} \mid\right.$ dismiss, $\left.\hat{\gamma}\right]=(1-\theta) \hat{\gamma}\left(s_{1}+s_{2}\right)$,
$E\left[L_{P} \mid\right.$ gototrial, $\left.\hat{\gamma}\right]=\hat{\gamma}(1-q(G))(1-\theta)\left(s_{1}+s_{2}\right)+(1-\hat{\gamma}) q(I) \theta\left(s_{1}+s_{2}\right)$.
It is easy to see that $E\left[L_{P} \mid\right.$ dismiss, $\left.\hat{\gamma}=\theta\right]>E\left[L_{P} \mid\right.$ gototrial, $\left.\hat{\gamma}=\theta\right]$, implying that $P$ will strictly prefer going to trial if $\hat{\gamma}=\theta$. In this case, it is credible for $P$ to proceed to trial.

Bayesian updating of posterior belief implies that:
$\hat{\gamma}=\frac{\gamma\left(1-\alpha_{i}\right)}{1-\gamma+\gamma\left(1-\alpha_{i}\right)}$.
From Eqs. (4) and (5), the equilibrium acceptance probability is $\alpha_{i}^{*}=$ $1-\frac{(1-\gamma) \theta}{\gamma(1-\theta)} \equiv \alpha^{*}$, for $i=1,2$.

Finally, consider the acceptance decisions of the defendants. Facing the plea offer $b_{i}$, defendant $D_{i}(t)$ will accept if:
$E\left[W_{i}\left(b_{i}\right) ;\right.$ accept $]=b_{i} \leq \beta_{i}\left[\alpha_{j}+\left(1-\alpha_{j}\right) q(t)\right] s_{i}=E\left(W_{i}\left(b_{i}\right) ;\right.$ reject $]$,
with indifference if this weak inequality holds with equality. If the other defendant $\left(D_{j}\right)$ accepts his plea offer with probability $\alpha_{j}$, then $P$ 's winning probability against $D_{i}$ becomes one and $D_{i}$ 's expected loss is $s_{i}$. If $D_{j}$ rejects his offer, then $D_{i}$ 's winning probability is $q(t)$ and expected loss is $q(t) s_{i}$. From the indifference condition for guilty defendants, it follows that:
$\beta_{i}=\frac{b_{i}}{\left[\alpha_{j}+\left(1-\alpha_{j}\right) q(G)\right] s_{i}}$.
From Eqs. (4)-(6), the equilibrium values of $\alpha_{i}^{*}, \beta_{i}^{*}$ and $\hat{\gamma}$ can be computed. Note that innocent defendants strictly prefer to reject their offers because $q(I)<q(G)$.

### 5.2. Possibility of fabricating evidence

So far, we assumed that if only one defendant accepts the plea offer and testifies, then the other defendant is convicted with probability one-even if both defendants are innocent. This assumption follows from the possibility that a defendant can fabricate evidence against his co-defendant. If evidence cannot be easily fabricated,


Fig. 7. Separating equilibrium when $r(G)<1$.
then it is more reasonable to assume that the conviction probability increases after the testimony of the co-defendant but remains uncertain. To capture this idea (allowing for a strictly positive chance that $P$ fails at trial even when one co-defendant has accepted a plea and provided evidence against the other), the analysis in this subsection assumes that $0<q(I)<r(I)<q(G)<r(G)<1$, where $r(t)$ denotes the conviction probability of type $t$ after the co-defendant's testimony who accepted a plea has been provided. After $P$ has offered $b_{1}$ and $b_{2}$, the defendants' decisions about whether to accept or reject may be affected by the modified assumption as follows:

$$
\begin{aligned}
& A A(t)=\left\{\mathbf{b}=\left(b_{1}, b_{2}\right) \in B \mid b_{i} \leq s_{i}, i=1,2\right\}, \\
& R R(t)=\left\{\mathbf{b} \in B \mid b_{i}>q(t) s_{i}, i=1,2\right\}, \\
& A R(t)=\left\{\mathbf{b} \in B \mid b_{1} \leq q(t) s_{1}, b_{2}>r(t) s_{2}\right\}, \\
& R A(t)=\left\{\mathbf{b} \in B \mid, b_{1}>r(t) s_{1}, b_{2} \leq q(t) s_{2}\right\} .
\end{aligned}
$$

Fig. 7 shows how the modified assumption of uncertain conviction following acceptance of one plea offer affects the four regions shown previously in Fig. 2. 25 Similar arguments lead us to conclude that the separating plea offers are $\left(r(G) s_{1}, q(G) s_{2}\right)$ while the pooling offers remain unaffected. Intuitively, $q(G) s_{2}$ is the greatest offer that the less culpable guilty defendant is willing to accept. $D_{1}$ knows that the less culpable defendant $D_{2}$ will receive such a plea discount and will accept it, leading to $D_{1}$ being convicted with probability $r(G)$. Therefore, $P$ will make a plea offer equal to the certainty equivalent of his trial outcome, $r(G) s_{1}$. Since the equilibrium outcome does not rely on $r(I)$, the analysis under this alternative assumption is almost the same as long as $r(G) \approx 1$.

### 5.3. Trial costs

It is widely recognized that plea bargaining has the potential to save the trial costs that would otherwise occur by going to trial. These costs are borne both by the prosecutor (and taxpayers he represents) and defendants. In this subsection, we allow for positive trial costs $C_{p}$ and $C_{d}$, representing trial costs to the prosecutor and to defendants, respectively, with $C_{p}>0$ and $C_{d}>0$.

We assume that $D_{1}$ and $D_{2}$ care about the trial costs as well as the penalty:
$\tilde{W}_{i}\left(x_{i} ; \tilde{x}_{i}\right)=x_{i}+\vartheta_{d} C_{d}$,

( $b_{1}$ )

Fig. 8. Separating equilibrium when $C d^{\prime}>0$.
where $\vartheta_{d}(>0)$ represents a defendant's marginal disutility of trial costs. Then, the four regions characterizing defendants' decisions regarding acceptance of plea offers are modified as follows:

$$
\begin{aligned}
& A A(t)=\left\{\mathbf{b}=\left(b_{1}, b_{2}\right) \in B \mid b_{i} \leq s_{i}+C_{d}^{\prime}, i=1,2\right\} \\
& R R(t)=\left\{\mathbf{b} \in B \mid b_{i}>q(t) s_{i}+C_{d}^{\prime}, i=1,2\right\} \\
& A R(t)=\left\{\mathbf{b} \in B \mid b_{1} \leq q(t) s_{1}+C_{d}^{\prime}, b_{2}>s_{2}+C_{d}^{\prime}\right\} \\
& R A(t)=\left\{\mathbf{b} \in B \mid, b_{1}>s_{1}+C_{d}^{\prime}, b_{2} \leq q(t) s_{2}+C_{d}^{\prime}\right\}
\end{aligned}
$$

where $C_{d}^{\prime}=\vartheta_{d} C_{d}$. All the boundaries shown earlier in Fig. 2 are shifted upward and to the right by $C_{d}^{\prime}$ as shown in Fig. 8. The result of this shift is that the region $A A$ is enlarged (i.e., defendants are more likely to accept plea offers). Intuitively, defendants know that losses from going to trial are greater by $C_{d}$ if they reject the prosecutor's plea offer. Fig. 8 shows the unique separating equilibrium offers are $\left(s_{1}, q(G) s_{2}+C_{d}^{\prime}\right)$. Because $D_{2}(G)$ accepts a larger offer (i.e., more severe punishment which is increased by $C_{d}^{\prime}$ ), typeII errors are reduced in this equilibrium, as long as $C_{d}^{\prime}$ is not too high $\left(q(G) s_{2}+C_{d}^{\prime}<s_{2}\right.$, or equivalently, $\left.C_{d}^{\prime}<(1-q(G)) s_{2}\right)$.

Next we consider how the prosecutor's decision is affected when she must incur trial cost $C_{p}>0$. In a separating equilibrium, guilty defendants accept plea offers and innocent defendants reject them, which lead to trial costs. In a pooling equilibrium, however, both types of defendants accept the pooling offers, which implies zero trial costs. Facing a positive trial cost, the prosecutor will therefore tend to prefer pooling offers over separating offers. The separating region becomes smaller (i.e., $\bar{\Theta}$ becomes smaller). For example, if $C_{d}=C_{p}=0.2$ and $\vartheta_{d}=\vartheta_{p}=1$, we obtain the new threshold value $\bar{\Theta}^{\prime} \approx$ $1.275<\bar{\Theta}=1.466$.

## 6. Incentive to commit a crime

The analysis above assumes that the defendants' decision of whether to commit a crime has already been made. In this section, we consider an extended model in which defendants decide whether to commit a crime, anticipating that plea bargaining will follow in the event they are prosecuted.

The extended game goes as follows. First, potential criminals, who can be viewed as potential defendants, $D_{1}$ and $D_{2}$ are informed of their respective private benefits from a crime, which will be denoted by $v_{1}$ and $v_{2}$. Then, they decide to commit a crime independently. If both decide to do so, the crime is committed. So, our implicit assumptions are: (i) that two potential defendants are partners, so that both of them are essential to committing the crime ${ }^{26}$;

[^11][^12]

Fig. 9. The sequence of events.
and (ii) their private benefits and costs are non-transferable so that neither one can compensate the other. If they are arrested with the probability $p, P$ makes plea offers $b_{1}$ and $b_{2}$ to them simultaneously without being informed of their guilt or innocence, and then $D_{1}$ and $D_{2}$ decide whether to accept their respective plea offer and if a defendant rejects it, he is proceeded to trial. See Fig. 9 for the timeline.

We assume that $v_{i}$ is distributed according to the distribution function $F\left(v_{i}\right)$ over $[0, \infty)$. For now, we assume that $p$ does not depend on whether the defendant is guilty or innocent; in other words, it is determined by circumstantial evidence without any direct evidence. ${ }^{27}$ Then, the potential defendant-to-be will choose to commit the crime whenever the net private benefit from committing the crime exceeds that individual's net private benefit from not committing the crime, which translates into the inequality $v_{i} \geq$ $p\left[E\left(\Delta W_{i}\right)\right]$, where $\Delta W_{i}=W_{i}^{C}-W_{i}^{N}$, and $W_{i}^{C}\left(W_{i}^{N}\right.$, respectively) represents $D_{i}$ 's loss from committing the crime (not committing the crime).

Let $x_{i}$ represent $D_{i}$ 's sure sentence. Because the two codefendants are assumed to both be involved, the probability that they jointly committed the crime is specified as $\gamma=\operatorname{Prob}\left(v_{1} \geq\right.$ $\left.p \Delta W_{1}\left(x_{1}\right), v_{2} \geq p \Delta W_{2}\left(x_{2}\right)\right)=\left[1-F\left(p x_{1}\right)\right]\left[1-F\left(p x_{2}\right)\right]$, assuming that no plea bargaining is allowed and no uncertainty is involved in the judicial decision. Thus, in the extended model, the probability that the two co-defendants are guilty is no longer fixed, as they were previously, but rather endogenously determined based on defendants' private benefits and costs, which lead to their respective decisions about whether or not to commit the crime. Clearly, we have $\partial \gamma / \partial x_{i}<0$ for $i=1,2$.

The effect of introducing plea bargaining on the crime rate will depend on whether the resulting value of $\Theta \equiv \frac{(1-\gamma) \theta}{\gamma(1-\theta)}$ yields a separating equilibrium or a pooling equilibrium in the subsequent plea bargaining game, since defendants make criminal decisions by anticipating the plea bargaining outcome. It is easy to see, however, that no pooling equilibrium exists in this extended game with given $p$, based on the following argument.

Potential defendants decide whether to commit a crime by comparing the net gains from committing the crime versus not committing the crime. Because the pooling plea offer to $D_{i}$ is $\Gamma s_{i}$ in an interior pooling equilibrium, the defendant's net gain from committing the crime is $v_{i}-p \Gamma s_{i}$, and his net gain from not committing a crime is $-p \Gamma s_{i}$, because he accepts the same plea offer $\Gamma s_{i}$ in a pooling equilibrium regardless of whether he commits the crime. Thus, they commit the crime if $v_{i} \geq p\left(\Gamma s_{i}-\Gamma s_{i}\right)=0$. In other words, potential defendants will always choose to commit the crime in a pooling equilibrium. The reason is that the opportunity cost of committing a crime is zero in a pooling equilibrium because the potential criminal's cost would be the same even if he did not commit the crime. The possibility of plea bargaining causes the opportunity cost of committing the crime to shift to zero in any pooling equilibrium. Therefore, potential defendants prefer to com-

[^13]crime rate function


Fig. 10. Determination of the pooling-equilibrium crime rate, $\gamma^{P}$, relative to the minimum crime rate for which separating equilibrium is supported, $\gamma$.
mit the crime in any pooling equilibrium (unless the private benefit from a crime is negative). Because defendants always commit the crime, then $\gamma=1,{ }^{28}$ and thus $\Gamma=1>\underline{\Gamma}(>0)$, which contradicts the condition for existence of the pooling equilibrium.

The main insight behind the result of no pooling in equilibrium is that although defendants have a higher criminal incentive (thereby inducing higher $\gamma$ ) due to lower opportunity costs in a pooling equilibrium than in a separating equilibrium, it contradicts with the interest of $P$, since $P$ prefers a pooling equilibrium to a separating equilibrium only if $\Theta$ is high, or equivalently, $\gamma$ is low. This insight is carried over to the more general case that $p_{C}>p_{N}(>0)$ where $p_{C}$ is the arrest probability when a defendant commits a crime and $p_{N}$ is the probability of wrongful arrest (when he does not commit a crime), as long as $\Delta p \equiv p_{C}-p_{N}$ is small. ${ }^{29}$

Proposition 3. There exists $\rho(>0)$ such that for any $\Delta p \leq \rho$, a pooling equilibrium does not exist in the extended game.

If $\Delta p$ is large (the police's arrest technology is sufficiently efficient), then a pooling equilibrium may be viable. Let the crime rate in an interior pooling equilibrium be $\gamma^{P}$. Then, it must satisfy
$\gamma=\left[1-F\left(\Delta p \Gamma(\gamma) s_{1}\right)\right]\left[1-F\left(\Delta p \Gamma(\gamma) s_{2}\right)\right] \equiv \Phi(\gamma)$,
where $\Phi(\gamma)$ represents the crime rate function in units of probability. Note that the pooling offer $\Gamma(\gamma) s_{i}$ increases in $\gamma$ and so does $F\left(\Delta p \Gamma(\gamma) s_{i}\right)$, implying that the right-hand side of Eq. (7) decreases in $\gamma$. Intuitively, as the crime rate $\gamma$ becomes larger, defendants are more likely to be guilty and thus type-II errors become more important. Therefore, the pooling offer $b_{i}$ tends to be higher, so the likelihood that the gain from a crime $\left(v_{i}\right)$ exceeds the opportunity cost of committing a crime ( $\Delta p b_{i}$ ) is lower. Fig. 10 shows how the fixed point of $\Phi($.$) function, \gamma^{P}$, is determined. If $\gamma^{P}<\gamma$, then $\gamma^{P}$ is a legitimate pooling equilibrium crime rate (and $\gamma$, as defined above, denotes the minimum value of $\gamma$ for which a separating equilibrium is supported). The crime rate in a non-interior pooling equilibrium can be similarly determined.

[^14]Comparison of the crime rate. We will compare the crime rate associated with plea bargaining versus without. Intuitively, plea bargaining has two conflicting effects on the incentive to commit a crime. On one hand, plea bargaining can increase the incentive to commit a crime because reduced sentences in plea offers consequently reduce the expected cost of committing a crime. On the other hand, plea bargaining may reduce the incentive to commit a crime because plea bargaining can improve informational efficiency, thereby increasing the probability that guilt is revealed. We will refer to the crime-incentivizing effect as the low-penalty effect and the crime-disincentivizing effect as the high-conviction-rate effect.

First, we will consider the case that the plea bargaining outcome is a separating equilibrium. It is not difficult to see that the separating equilibrium obtained in Proposition 1 remains unaffected in this extended game. Given the criminal decision being made, $P$ has no reason to make a different pair of plea offer because it cannot affect the criminal incentive. Now, note that the outcome in the case of innocent defendants is the same regardless of whether plea bargaining is used, because innocent defendants will reject plea offers in the separating equilibrium. In the case of guilty defendants, the less culpable defendant faces exactly the same incentive to commit the crime because his penalty under plea bargaining is the certainty equivalent, $q(G) s_{2}$, of his uncertain penalty when plea bargaining is not allowed (i.e., he gets the sentence $s_{2}$ with probability $q(G)$ ). In contrast, the incentive of the more culpable defendant changes once plea bargaining is introduced. Under plea bargaining, the more culpable defendant receives the plea offer $s_{1}$ and chooses to accept it. Therefore, his expected penalty under plea bargaining, $s_{1}$, must be greater than the expected penalty without plea bargaining, $q(G) s_{1}$, because, under plea bargaining, the more culpable defendant is subject to the same penalty but with a higher conviction probability due to the other defendant's testimony, which is precisely the high-conviction effect. Because of this difference, plea bargaining has the effect of reducing the crime rate in a separating equilibrium by increasing the probability of conviction, i.e., $\gamma^{S}<\gamma^{N}$ where $\gamma^{S}$ and $\gamma^{N}$ are crime rates under the plea-bargaining (in a separating equilibrium) and no-plea-bargaining institutions, respectively.

The reason why the low-penalty effect does not appear in this separating equilibrium is that the plea offer made to the less culpable defendant is not just a reduced penalty but a rather elaborately calculated penalty which turns out to be exactly the same as the expected penalty (i.e., certainty equivalent) without plea bargaining. By making such an offer (which is not lower than the expected penalty $q(G) s_{2}$ ) and taking the possibility of type-II error into account, $P$ can avoid weakening the deterrent effect.

We turn now to the case that the plea bargaining game selects a pooling equilibrium. The viability of a pooling equilibrium requires that $\gamma^{P}<\underline{\gamma}$. Also, in any separating equilibrium, we must have $\underline{\gamma}<$ $\gamma^{S}<\gamma^{N}$. Therefore, we can conclude that $\gamma^{P}<\gamma^{N}$, which means that the crime rate is lowered by plea bargaining even in a pooling equilibrium. The intuition is rather tricky, because it is not immediately clear how to compare the severity of expected penalties in the two cases of the no-plea-bargaining outcome and the pooling equilibrium with plea bargaining. From the static model, it seems that pooling offers are much less severe than equilibrium separating offers are and also less severe than expected penalties without plea bargaining. While it is true that separating-equilibrium plea offers are more severe than pooling-equilibrium plea offers for any given $\gamma$, the two cases of separating and pooling offers never arise for a single value of $\gamma$. Pooling equilibria are chosen only for sufficiently low values of $\gamma$. In other words, $P$ does not choose to make pooling offers whenever there are concerns that they will lead to a higher crime rate. When can $P$ expect that the result-
ing crime rate will be sufficiently low, satisfying the condition $\gamma^{P}<\underline{\gamma}$ ? $P$ will prefer a pooling equilibrium to a separating equilibrium if she is very unlikely to win the case by going to court, (i.e., if $q(G)$ is low, satisfying $q(G)<\Gamma$ ). In this case, how much will the defendants expect their penalties to be reduced if they choose not to commit a crime? The reduction in the expected penalties is larger under plea bargaining (in a pooling equilibrium) than under no plea bargaining, because $\Gamma\left(p_{G}-p_{I}\right)>p_{G} q(G)-p_{I} q(I)$ or equivalently, $[\Gamma-q(G)] p_{G}>[\Gamma-q(I)] p_{I}$ if $p_{G} \gg p_{I}$. This implies that if $p_{G}$ and $p_{I}$ differ by much, defendants have a less incentive to commit a crime under plea bargaining even in the case of the pooling equilibrium. Again, the intuitive reason why a harsher sentence is feasible with plea bargaining in a pooling equilibrium than without plea bargaining is that $P$ can exploit the informational advantage that follows from the testimony of the other defendant (if a defendant rejects the plea offer) as a threat. The following proposition summarizes the analysis presented in this section.

## Proposition 4. Plea bargaining reduces the crime rate.

Comparison under P' alternative objective function. To see the robustness of Proposition 4, we will consider the possibility of alternative objective functions for the prosecutor. The prosecutor may, for example, be concerned about the social harm caused by the crime as well as by judicial errors. If we consider this possibility as well, then the corresponding loss function $\hat{L}_{P}$ would be a weighted sum of judicial errors and social harm caused directly by the crime:
$\hat{L}_{P}\left(x_{1}, x_{2}\right)=\lambda \sum_{i=1}^{2} L_{i}\left(x_{i}\right)+(1-\lambda) \gamma\left(x_{1}, x_{2}\right) H$,
where $H$ is the social harm from the crime itself and $\lambda$ is the relative weight placed on judicial errors. Note that the second term (social harms) as well as the first term (judicial errors) are affected by the actual expected sentences $x_{1}$ and $x_{2}$ because the crime rate is specified to depend on $x_{1}$ and $x_{2}$.

However, even if we consider this alternative loss function of the prosecutor, her decision is not affected in a separating equilibrium. To see this, consider once again point $E_{1}$ in Fig. 4, whose coordinates provide the unique separating equilibrium offers in the plea-bargaining game under the objective function $L_{P}$. We will show that the point is still the unique equilibrium outcome under $\hat{L}_{P}$. In a separating equilibrium, if we want to stay within the yellow region labeled $K=A A(G) \cap R R(I) \subset B^{S}$ in Fig. 4 so as to slightly reduce $x_{1}$ or $x_{2}$ (within that region), then the overall crime rate increases. More generally, let the function $\gamma=\left[1-F\left(p x_{1}\right)\right]\left[1-F\left(p x_{2}\right)\right] \equiv g\left(x_{1}\right.$, $x_{2}$ ) represent the indifference curve that yields the same crime rate. Then, the indifference curve is downward-sloping and symmetric along $x_{1}=x_{2}$, because $\partial g / \partial x_{i}<0$ for $i=1,2$ and $F\left(v_{i}\right)$ is identical for $i=1,2$. The upper contour set of crime rate $\gamma_{0}$ is defined by $U=\left\{\left(x_{1}\right.\right.$, $\left.\left.x_{2}\right) \mid \gamma_{0} \leq g\left(x_{1}, x_{2}\right)\right\}$. If we let $\gamma_{0}=g\left(s_{1}, q(G) s_{2}\right)$, any $\left(b_{1}, b_{2}\right) \in K$ must be in $U$, implying that it increases the crime rate. Therefore, social loss cannot be reduced by such a change. In the opposite direction, if we consider the region $K^{C} \cap B^{S}$ (e.g., point $A^{\prime}$ in Fig. 4) hoping to reduce the crime rate, then this offer is necessarily rejected by one of the defendants (say $D_{1}$ ), and the actual expected penalty is reduced to ( $s_{1}, q(G) s_{2}$ ). Because this outcome is identical to the plea offer $\left(b_{1}, b_{2}\right)=\left(s_{1}, q(G) s_{2}\right)$ in the view of defendants, it will leave the crime rate unaffected. Therefore, the unique separating equilibrium obtained in Proposition 1 remains the unique separating equilibrium in the extended game under the objective function $\hat{L}_{P} .{ }^{30}$

[^15]Pooling equilibria could be affected by the alternative objective function of the prosecutor based on its implicit extra consideration of crime deterrence. If the prosecutor is concerned about the deterrence effect, she will increase her pooling offers to sacrifice judicial errors by moving northeast slightly from the static equilibrium, which will definitely lead to a lower crime rate. Therefore, the result that plea bargaining will reduce a crime rate is not affected under this alternative objective function of the prosecutor; rather, it is strengthened. ${ }^{31}$ Note that this result does not rely on P's ability to commit to some $b_{1}$ and $b_{2}$ in the plea bargaining stage. Since the potential defendants can rationally anticipate how $b_{1}$ and $b_{2}$ will be determined in equilibrium, they will be less inclined to commit a crime (without commitment) if $P$ has a harm-added objective function.

Welfare comparison. Next, we examine whether introducing plea bargaining is socially beneficial. Intuitively, plea bargaining has the additional effect on social welfare of reducing judicial errors as well as its two effects on the crime rate. Since social welfare is determined by judicial errors and the crime rate, we can define the social loss function $\left(L_{S}\right)$ which is identical to the prosecutor's alternative loss function $\left(\hat{L}_{P}\right)$ : that is, $L_{S}=\lambda \sum_{i=1}^{2} L_{i}\left(x_{i}\right)+(1-\lambda) \gamma\left(x_{1}, x_{2}\right) H$. To compare the social losses when plea bargaining is allowed versus not allowed, it is convenient to first compare the social losses due to judicial errors (i.e., the first term in $L_{S}$ ).

First, we will compare the no-plea-bargaining outcome with plea bargaining when the subsequent equilibrium is separating. Equilibrium social losses (due to judicial errors) with no plea bargaining versus with plea bargaining in a separating equilibrium are denoted as $L^{N}$ and $L^{S}$, respectively, and computed as follows:
$L^{N}=\gamma^{N}(1-\theta)(1-q(G))\left(s_{1}^{2}+s_{2}^{2}\right)+\left(1-\gamma^{N}\right) \theta q(I)\left(s_{1}^{2}+s_{2}^{2}\right)$,
$L^{S}=\gamma^{S}(1-\theta)(1-q(G))^{2} s_{2}^{2}+\left(1-\gamma^{S}\right) \theta q(I)\left(s_{1}^{2}+s_{2}^{2}\right)$.
As argued earlier, the introduction of plea bargaining has two advantages reflected in the social loss function. First, plea bargaining lowers the crime rate $\left(\gamma^{S}<\gamma^{N}\right)$. Second, the risk of type-II error is reduced: $\left((1-q(G))^{2} s_{2}^{2}<(1-q(G))\left(s_{1}^{2}+s_{2}^{2}\right)\right)$. The second effect is due to $P$ 's quadratic loss function that represents prosecutorial risk aversion, given that the actual penalty that guilty defendants face under plea bargaining is the certainty equivalent of their risky penalty under no plea bargaining. Although both of the effects are advantageous to social welfare, combining the two effects may, paradoxically, lower social welfare, because the first effect (a decrease in $\gamma$ ) makes the second effect less important $\left(1-\gamma^{S}>1-\gamma^{N}\right)$. Because the defendants are more likely to be innocent under plea bargaining, it may decrease social welfare if the risk of type-I error is greater than that of type-II error $\left((1-\theta)(1-q(G))^{2} s_{2}^{2}<q(I)\left(s_{1}^{2}+s_{2}^{2}\right)\right)$. However, if we assume that the wrongful conviction probability of innocent defendants $q(I)$ (the main source of type-I error) is very low, then we can say that plea bargaining is socially beneficial i.e., $L^{S}<L^{N}$, because its first-order effects are to deter crime and reduce type-II errors.

If a pooling equilibrium occurs under the plea bargaining system, we should compare the equilibrium social losses denoted by $L^{P}$ with $L^{N}$. However, if $P$ prefers a pooling equilibrium to a separating equilibrium, it means that $L^{P}<L^{S}$, in turn, implying that
needs to be concerned about judicial errors and ignore any effects of plea bargaining on incentives of potential criminals to commit crimes (because the crime rate is already determined, i.e., the prosecutor's loss function is now reduced to $L_{P}$ ). However, because minimizing $\hat{L}_{P}$ is equivalent to minimizing $L_{P}$, then no problem of time inconsistency occurs.
${ }^{31}$ The result of no pooling equilibrium when $\Delta p$ is small is also valid regardless of the loss function for $P$ (i.e., either $L_{P}$ or $\hat{L}_{P}$ ), because $\gamma=1$ is not consistent with the viability of the pooling equilibrium.
$L^{P}<L^{N}$. Therefore, plea bargaining increases social welfare whether a separating equilibrium or a pooling equilibrium results.

Finally, we consider the entire social loss function $L_{S}$ and easily see that the social efficiency of the plea bargaining institution remains. If we add the term $\gamma H$ to the social loss function due to judicial errors $\left(L_{P}\right)$, then the social efficiency of plea bargaining is strengthened because of its additional effect of reducing social losses by decreasing the crime rate. Counter to widespread perceptions to the contrary, the case for plea bargaining is strengthened by taking into account the social harms caused by the crime itself, at least in the multi-defendant setting.

## 7. Conclusion and caveats

In this article, we considered the incentives to commit crimes among multiple potential co-defendants and some new issues surrounding asymmetric information between prosecutor and defendants in a model of plea bargaining with multiple defendants.

One intriguing direction for extending our model would be to incorporate the possibility that the two defendants' types are not perfectly correlated, that is, that the prosecutor is not sure about whether one defendant is a sole criminal or has an accomplice. Extending to those cases by allowing the possibility that one defendant is innocent while the other is guilty, the model's implications based on the mechanism of inferring one defendant's guilt from the other's guilty plea would likely require substantial modifications.

## Appendix.

## Proof of Proposition 1

Proof of Proposition 1. Since $L_{P}(\mathbf{b})$ is monotonically decreasing in both $b_{1}$ and $b_{2}$, the equilibrium offers must be either $\mathbf{b}_{1}=\left(s_{1}\right.$, $\left.q(G) s_{2}\right)$ at $E_{1}$ or $\mathbf{b}_{2}=\left(q(G) s_{1}, s_{2}\right)$ at $E_{2}$. We will compare expected losses at $E_{1}$ and at $E_{2}$. We have
$L_{P}\left(\mathbf{b}_{1}\right)=(1-\gamma) \theta q(I)\left(s_{1}^{2}+s_{2}^{2}\right)+\gamma(1-\theta)(1-q(G))^{2} s_{2}^{2}$,
$L_{P}\left(\mathbf{b}_{2}\right)=(1-\gamma) \theta q(I)\left(s_{1}^{2}+s_{2}^{2}\right)+\gamma(1-\theta)(1-q(G))^{2} s_{1}^{2}$.
Since type-I errors are the same at $E_{1}$ and $E_{2}$, we have $L_{P}\left(\mathbf{b}_{1}\right)<L_{P}\left(\mathbf{b}_{2}\right)$ if and only if $s_{1}>s_{2}$.
Proof of Lemma 2. The proof follows from comparing P's losses from the offers in the pooling equilibrium and his losses from the separating equilibrium offers. When $b_{i}>s_{i}$ for $i=1,2, P$ 's losses are:
$L_{P}\left(b_{1}, b_{2}\right)=\gamma(1-\theta)(1-q(G))\left(s_{1}^{2}+s_{2}^{2}\right)+L_{0}$,
where $L_{0}=(1-\gamma) \theta q(I)\left(s_{1}^{2}+s_{2}^{2}\right)$, because both offers are always rejected. On the other hand, $P$ 's losses from the separating offers are:
$L_{P}\left(s_{1}, q(G) s_{2}\right)=\gamma(1-\theta)\left(s_{2}-q(G) s_{2}\right)^{2}+L_{0}$.
One can easily see that:
$L_{P}\left(s_{1}, q(G) s_{2}\right)<L_{P}\left(b_{1}, b_{2}\right) \Leftrightarrow(1-q) s_{2}^{2}<s_{1}^{2}+s_{2}^{2}$,
which always holds. Thus, the proof is complete. $\square$
Proof of Lemma 2. Consider an interior point in $A A(I)$. Note that both offers at this point are accepted by both types of defendants. Then any point except points on the red line segment is associated with a lens-shaped region (as illustrated in Fig. 5) generated by two iso-loss circles passing through the point so that a deviation to a point in the lens area reduces both type-I and type-II errors. This result implies that all the possible pooling offers in the interior of $A A(I)$ region are in the red line segment. $\square$

Proof of Lemma 3. Consider the vertical boundary of $A A(I)$. Any point except the one point on the red line segment is, similar as in the previous paragraph, associated with a loss-reducing lensshaped area overlapping with $A A(I)$, so that it cannot be a pair of pooling equilibrium offers. On the other hand, the lens area associated with a point on the horizontal boundary of $A A(I)$ does not overlap with $A A(I)$, which means that $P$ cannot choose any better pooling offers by deviating to a point in the loss-reducing lens area.

Proof of Proposition 2. The pooling region consists of $A A(I), A R(I)$, $R A(I)$ and $R R(G)$, but we omit considering $R R(G)$ due to Lemma 1 .

Consider any point in $A R(I) \cup R A(I)$. Because a pair of offers in $A R(I)$ and $R A(I)$ is rejected by $D_{2}(I)$ and $D_{1}(I)$, respectively, the pair of sentences that the defendants actually expect to receive will end up on the border between $A A(I)$ and $A R(I)$ (or $R A(I)$, respectively). As before, $P$ can find a better pair of offers than any point in the border between $A A(I)$ and $A R(I)$ by moving inside a loss-reducing lens area, except for the point $\left(q(I) s_{1}, s_{2}\right)$. Similarly, $P$ can profitably deviate from any point in the border between $A A(I)$ and $R A(I)$ except for the point ( $s_{1}, q(I) s_{2}$ ).

Regarding point $\left(q(I) s_{1}, s_{2}\right)$, one can see that the intersection of the loss-reducing lens area and the set $A A(I)$ is not empty as long as $s_{1}>s_{2}$. This means that $P$ can profitably deviate from the point, which implies that point $\left(q(I) s_{1}, s_{2}\right)$ cannot be a pair of pooling equilibrium offers. At point $\left(s_{1}, q(I) s_{2}\right)$, there is no intersection of the lens area and the set $A A(I)$. Thus, the result follows from Lemmas 2 and 3 .

Proof of Proposition 3. If $p_{C}=p_{N}=p$, in a pooling equilibrium, a potential defendant commits a crime if $v_{i} \geq p\left(\Gamma s_{i}-\Gamma s_{i}\right)=$ 0 . Because $F(0)=0, \gamma\left(b_{1}, b_{2}\right)=1-F(0)=1$ for any pooling offers ( $b_{1}, b_{2}$ ), implying that $\Gamma=1$. Because $P$ prefers pooling offers to separating offers only if $\Gamma \leq \underline{\Gamma}(<1)$, it is a contradiction. If $\Delta p=p_{C}-p_{N}>0, \lim _{\Delta p \rightarrow 0} \gamma(\Delta p)=\lim _{\Delta p \rightarrow 0}\left[1-F\left(b_{1} \Delta p\right)\right]\left[1-F\left(b_{2} \Delta p\right)\right]=$ 1. Therefore, there exists $\rho(>0)$ such that for any $\Delta p<\rho, \gamma(\Delta p)>\gamma$, which violates the condition for the existence of a pooling equilibrium.

Proof of Proposition 4. $\quad \gamma^{S}=\gamma\left(s_{1}, q(G) s_{2}\right)<\gamma\left(q(G) s_{1}, q(G) s_{2}\right)=\gamma^{N}$ because $\partial \gamma / \partial x_{i}<0$. Because $\gamma^{P}<\underline{\gamma}<\gamma^{S}<\gamma^{N}$, the proof is complete.

## References

Baker, S., Mezzetti, C., 2001. Prosecutorial resources, plea bargaining, and the decision to go to trial. J. Law Econ. Organ. 17, 149-167.
Bar-Gill, O., Ben-Shahar, O., 2009. The prisoners' (plea bargain) dilemma. J. Legal Anal. 1, 737-773.
Daughety, A., Reinganum, J., 2016. Informal sanctions on prosecutors and defendants and the disposition of criminal cases. J. Law Econ. Organ. 32, 359-394.
French Ministry of Justice, 2006. Les chiffres-cles de la Justice.
Goode, E., 2012. Stronger Hand for Judges in the 'Bazaar' of Plea Deals. New York Times, March 22, 2012.
Grossman, G., Katz, M., 1983. Plea bargaining and social welfare. Am. Econ. Rev. 73, 749-757.
Guidorizzi, D., 1998. Comment: Should we really "ban" plea bargaining?: the core concerns of plea bargaining critics. Emory Law J. 47, 753-783.
Harsanyi, J., Selten, R., 1988. A General Theory of Equilibrium Selection in Games. The MIT Press, Cambridge, MA.
Kim, J.-Y., 2009. Secrecy and fairness in plea bargaining with multiple defendants. J. Econ. 96, 263-276.

Kim, J.-Y., 2010. Credible plea bargaining. Eur. J. Law Econ. 29, 279-293.
Kobayashi, B., 1992. Deterrence with multiple defendants: an explanation for "unfair" plea bargains. RAND J. Econ. 23, 507-517.
Kornhauser, L., Revesz, R., 1994a. Multidefendant settlements: the impact of joint and several liability. J. Legal Stud. 23, 41-76.
Kornhauser, L., Revesz, R., 1994b. Multidefendant settlements under joint and several liability: the problem of insolvency. J. Legal Stud. 23, 517-542.
Landes, W., 1971. An economic analysis of the courts. J. Law Econ. 14, 61-107.
Miceli, T., 1996. Plea bargaining and deterrence: an institutional approach. Eur. J. Law Econ. 3, 249-264.
Reinganum, J., 1988. Plea bargaining and prosecutorial discretion. Am. Econ. Rev. 78, 713-728.
Reinganum, J., 1993. The law enforcement process and criminal choice. Int. Rev. Law Econ. 13, 115-134.
Silva, F., 2017. If We Confess Our Sins. Mimeo.
Spier, K., 1994. A note on joint and several liability: insolvency, settlement and incentives. J. Legal Stud. 23, 559-568.
Spier, K., 2002. Settlement with multiple plaintiffs: the role of insolvency. J. Law Econ. Org. 18, 295-323.
Tor, A., Gazal-Ayal, O., Garcia, S., 2010. Fairness and the willingness to accept plea bargain offers. J. Empir. Legal Stud. 7, 97-116.


[^0]:    We are grateful to seminar participants at University of Kyushu and the Korean Foundation of Advanced Studies (KFAS), and audiences at the annual conference of the Korean Law and Economics association held in Seoul National University and the annual conference of the Asian law and economics association held in Academia Sinica, Taiwan and the conference on the anti-trust law sponsored by FTC and LEG Working Group for valuable comments. We gratefully acknowledge the Australian Research Council for financial support, Project DP150100242.

    * Corresponding author.

    E-mail addresses: Nathan.berg@otago.ac.nz (N. Berg), jyookim@khu.ac.kr (J.-Y. Kim).

[^1]:    ${ }^{1}$ Moreover, these papers do not consider defendants' private information about guilt or innocence as we do in our model. In the context of civil litigation, Spier (1994) considers a model of multiple defendants with incomplete information, while Kornhauser and Revesz (1994a,b) provide a complete information model of multiple defendants. While circulating this paper, we found Silva (2017) which addressed a similar issue. His paper differs from ours in two ways. First, he uses the mechanism design approach. Second, more importantly, he does not touch the dynamic issue of examining the incentive to commit a crime, which is the main issue of our paper.

[^2]:    ${ }^{2}$ http://www.americanbar.org/
    ${ }^{3}$ In a separating equilibrium, the plea bargain offers are asymmetric in that only less culpable defendants are offered plea discounts. In a pooling equilibrium, both defendants are offered plea discounts.
    ${ }^{4}$ Tor et al. (2010) show empirically that unfair plea offers are very likely to be rejected, which would seem to suggest that the fair plea outcome predicted by our model matches the available observational data.

[^3]:    ${ }^{5}$ If $s_{1}>s_{2}$, then $D_{1}$ and $D_{2}$ could be interpreted as principal and accessory, respectively; whereas if $s_{1}=s_{2}$, then $D_{1}$ and $D_{2}$ could be interpreted as principal and accomplice. In most countries' criminal law, an accomplice is subject to the same criminal penalties as the principal offender.
    ${ }^{6}$ We acknowledge there is a potentially important distinction between the concepts of guilt and culpability. The degree or intensity of guilt (i.e., "guiltiness") of a defendant concerns whether he committed a crime or not, which could be viewed as a binary state rather than a continuously valued parameter. In contrast, culpability is concerned with how severe the crime was and the extent to which the defendant is responsible. Thus, it is more reasonable to interpret $s_{i}$ as defendant $i$ 's culpability.
    ${ }^{7}$ Following Grossman and Katz, we are implicitly assuming that $P$ can effectively commit to not dismissing the case.
    ${ }^{8}$ Grossman and Katz (1983) and Reinganum (1988) also assume that a guilty defendant is more likely to be convicted than an innocent one.

[^4]:    ${ }^{9}$ An alternative assumption would be that the conviction probability when the other defendant accepts the plea offer is $r(t) \in(q(t), 1)$. This alternative would only change the size of equilibrium offers without affecting the qualitative features of equilibrium. Further detail is provided in Section 5.2.
    ${ }^{10}$ There is a rule (Rule 11) in the US Federal Rules of Criminal Procedure against admissibility of plea discussions as evidence in court.
    ${ }^{11}$ If $J$ realizes the possibility of his own judicial errors, then he or she may prefer a smaller sentence rather than $s_{i}$; but reduced sentences are not permitted under our assumption of no judicial discretion. Therefore, our setup rules out the possibility of intermediate judgements.
    ${ }^{12}$ In Section 5.3, we investigate the case of positive trial costs.
    ${ }^{13}$ We will simply refer to type-I errors and type-II errors instead of "losses due to type-I errors" and "losses due to type-II errors" whenever there is no chance of confusion.
    ${ }^{14}$ A function $\|\cdot, \cdot\|$ is called a metric if it satisfies (i) $\|x, y\| \geq 0$ with equality if and only if $x=y$, (ii) $\|x, y\|=\|y, x\|$ and (iii) $\|x, z\| \leq\|x, y\|+\|y, z\|$, for all $x, y, z$. In particular, if $\|x, y\|=|x-y|^{n}$, then $P$ is risk-neutral if $n=1$, risk-averse if $n>1$, and risk-loving if $n<1$.

[^5]:    ${ }^{15}$ Alternatively, a model in this context could instead assume that prosecutor and co-defendants conduct separate plea-offer negotiations in which defendant $D_{i}$ only observes the offer made to him alone, $b_{i}$. See Kim (2009) for the case of separate negotiations.
    ${ }^{16}$ We assume that any ties in payoffs where a player might otherwise be indifferent are resolved in favor of acceptance. This assumption is relaxed in Section 5.1.
    ${ }^{17}$ Kobayashi (1992) and Spier (2002) used the criterion of risk dominance instead of Pareto dominance. The risk dominance criterion, first proposed by Harsanyi and

[^6]:    Selten (1988), can be rationalized by the so-called linear tracing procedure. The main theoretical argument in favor of risk dominance is based on risky payoffs. However, co-defendants in our model have chances to agree on an outcome before they play the plea bargaining game. It is therefore more reasonable (i.e., more attractive and cognitively simpler) for them to focus on the outcome which is Pareto superior.
    ${ }^{18}$ Bar-Gill and Ben-Shahar (2009) consider an interesting case in which $P$ has enough resources to take only one defendant to trial. In this case, Fig. 2 can be modified by replacing $\left(q(t) s_{1}, q(t) s_{2}\right)$ with $\left(q(t) s_{1}, 0\right)$, since only the most culpable defendant stands trial if both defendants reject. This modified plea bargaining game is not a Prisoners' Dilemma, either, because the less culpable defendant prefers rejecting his plea offer if he believes that the more culpable defendant will reject his offer. The more culpable defendant will, however, never reject the offer, because it is his dominant strategy to accept the offer as long as he will be the one who stands trial if both defendants reject. So, both defendants will accept the offer. Therefore, the individually rational outcome under plea bargaining will deviate from the collectively rational one.

[^7]:    ${ }^{19}$ It is not difficult to see that $P$ is indifferent between $E_{1}$ and some pair of offers, one of which is rejected, say $A^{\prime}$. Then, $A^{\prime}$ is indeed a separating equilibrium pair of offers. But we can focus only on $E_{1}$ due to our assumption that indifference is resolved in favor of acceptance, as stated earlier in Footnote 16.

[^8]:    ${ }^{20}$ Indeed, this result hinges crucially on our assumption that the conviction probability after a guilty plea is $r(G)=1$. If $r(G)<1$, the equilibrium will require that $P$ discount the plea offer more to induce the more culpable defendant to accept it.
    21 http://www.justice.gov/atr/corporate-leniency-policy
    ${ }^{22}$ In fact, this is the unique equilibrium outcome in the game of complete information in which defendants are known to be guilty, which means that plea offers are fair even in the case of complete information.

[^9]:    ${ }^{23}$ The necessary and sufficient condition is that $\mathbf{b}^{*} \in \operatorname{argmin}_{\mathbf{b} \in B} L_{P}(\mathbf{b})=L_{1}(\mathbf{b})+L_{2}(\mathbf{b})$ and $\mathbf{b}^{*} \in B^{P}$.

[^10]:    ${ }^{24}$ See, for example, Reinganum (1988), who considered a model of two-sided uncertainty by assuming that the prosecutor also has private information. Her separating equilibrium (like ours) turns out to not be viable once she dispenses with the assumption that the prosecutor must go to trial after an offer is rejected.

[^11]:    ${ }^{26}$ The assumption of essential partners implies that each of them is nonsubstitutable in crime production, possibly due to their expertise.

[^12]:    25 Analysis in this subsection continues to rely on the Pareto dominance criterion for equilibrium selection.

[^13]:    ${ }^{27}$ For example, an innocent defendant may be arrested simply because he or she was at the crime scene together with the victim. Miceli (1996) also assumes that $p$ is exogenously given, whereas $p$ is endogenously determined in Reinganum (1993). Because our focus is the effect of plea bargaining on the crime rate, our modeling approach abstracts from the mechanisms that endogenously determine law-enforcement and law-making policies.

[^14]:    ${ }^{28}$ Note that this result of $\gamma=1$ comes from the assumption that $v_{i} \geq 0$. If there are some people who get direct utility from law abiding for some (moral) reasons, it may not be valid.
    ${ }^{29}$ To see this, note that $\gamma=\left[1-F\left(b_{1} \Delta p\right)\right]\left[1-F\left(b_{2} \Delta p\right)\right]$ and $\lim _{\Delta p \rightarrow 0} \gamma(\Delta p)=1$,
    because the pooling plea offers $b_{1}$ and $b_{2}$ are bounded above.

[^15]:    ${ }^{30}$ Uniqueness holds whether the prosecutor's loss function is $L_{P}$ or $\hat{L}_{P}$. It should be pointed out that the loss function $\hat{L}_{P}$ may raise the issue of time inconsistency. Once defendants have committed a crime, the prosecutor may decide that she only

