1 Introduction

A common criticism of plea bargaining is that the system frequently induces innocent defendants to plead guilty (Gazal-Ayal, 2005; Guidorizzi, 1998; Langbein, 1978). Prosecutors often offer large sentence reductions in return for guilty pleas (Bibas, 2004; Covey, 2007; McCoy, 2005). These sentence reductions are called “plea discounts” (Gazal-Ayal, 2005). Many worry that when plea discounts are large, there is enough incentive to settle that even innocent defendants plead guilty.

A common proposal to stop innocent defendants from pleading guilty is capping the size of plea discounts (Covey, 2007; Guidorizzi, 1998). Proponents of capped plea discounts argue that preventing innocent defendants from pleading guilty leaves innocent defendants better off. While this argument has received support within the legal community, it has encountered resistance within the economics community because it defies the traditional argument that voluntary contracting is Pareto efficient (Gazal-Ayal, 2005). Indeed, since a prosecutor and a defendant should only agree to a plea bargain when its terms are mutually beneficial, it seems counterintuitive that preventing a defendant from pleading guilty could help either party.

Several recent papers, however, have suggested that prosecutorial resource constraints complicate the welfare implications of plea bargaining reform (Bar-Gill and Ben-Shahar, 2009; Bar-Gill and Gazal-Ayal, 2006; Gazal-Ayal, 2005). In particular, Bar-Gill and Ben-Shahar (2009) note that when a prosecutor has limited resources, the decision of one defendant to plead guilty imposes a negative externality on other defendants by enabling the prosecutor to reallocate resources towards her other cases. Bar-Gill and Ben-Shahar suggest that while it may be in the individual self-interest of defendants to plea bargain, it is possible that defendants would be collectively better off if there were limits on bargaining.

Building upon this literature, this paper analyzes a multilateral model of plea bargaining under asymmetric information to investigate how prosecutorial resources impact
the welfare effects of capped plea discounts. I consider a game in which a prosecutor
offers take-it-or-leave-it pleas to two defendants before costly trial. While each defendant
knows whether he is innocent or guilty, the prosecutor knows only the joint probability
distribution of the defendants’ types. The types of the defendants are independent, and
guilty types are more likely to be convicted at trial. The core feature of the model is
the prosecutor’s trial capacity. I begin by supposing that the prosecutor is able to try
both defendants, and then suppose that the prosecutor only has the resources to try one
of the two defendants. The introduction of this resource constraint creates a negative
bargaining externality between the defendants, as the decision of one defendant to accept
his plea concentrates the trial threat on the other defendant. While this resource con-
straint was first considered by Bar-Gill and Ben-Shahar (2009), my paper is the first to
incorporate the constraint into a model of asymmetric information. More generally, my
paper presents one of the first multilateral models of plea bargaining with asymmetric
information.

I find that the introduction of the resource constraint reverses the effect that capped
plea discounts have on the welfare of innocent defendants. When the prosecutor is able
to try both defendants, capped plea discounts leave all parties weakly worse off. On the
other hand, when the prosecutor is only able to try one defendant, capped plea discounts
may help innocent defendants. Importantly, this improvement for innocent defendants
does not necessarily indicate an improvement for all defendants, as the reform may si-
multaneously decrease guilty defendant welfare and increase innocent defendant welfare.
These results are robust both when the prosecutor may commit ex ante to trial and
when offers are made sequentially rather than simultaneously. Given that prosecutors
have severely limited resources, I suggest that these findings not only illustrate the theo-
retical implications of limited prosecutorial resources for plea bargaining reform but also
provide practical economic support for capped plea discounts.

My paper has the following structure. Section 2 discusses related literature, summa-
rizing both the plea bargaining debate within the legal literature and relevant models of bargaining within the economics literature. Section 3 presents the multilateral model. Section 4 analyzes the model when the prosecutor is able to try both defendants. Section 5 analyzes the model when the prosecutor is able try only one of the two defendants. Section 6 demonstrates that the findings from Sections 4 and 5 are robust when plea offers are made sequentially rather than simultaneously. Section 7 concludes.
2 Related Literature

2.1 The Plea Bargaining Reform Debate

In the United States, 90 to 95 percent of all criminal convictions are obtained through plea bargains (U. S. Department of Justice, 2006; U. S. Sentencing Commission, 2010). This systematic reliance on plea bargains has generated fierce controversy, much of which has focused on the possibility that many of the defendants who plead guilty are in fact innocent (Gazal-Ayal, 2005; Guidorizzi, 1998; Langbein, 1978). This criticism of plea bargaining, often called plea bargaining’s “innocence problem,” has prompted many people to argue that reform is needed to protect innocent defendants. While some reformers advocate complete abolition of the practice (Alschuler, 1983; Schulhofer, 1992), most acknowledge that plea bargaining is necessary to preserve limited state resources (Alschuler, 1983; Bibas, 2004; Guidorizzi, 1998). Indeed, given that plea bargains save approximately 80 to 90 percent of the cost of trial, it seems unlikely that the legal system could handle its current case load without plea bargains (Schulhofer, 1988).^{1}

A commonly proposed solution to the innocence problem that has received attention within the legal community is to institute a cap on the size of plea discounts (Covey, 2007; Gazal-Ayal, 2005; Guidorizzi, 1998). While different authors advocate different variations of capped plea discounts,^{2} they agree that the essential feature of capped plea discounts is that they limit the incentive to plead guilty. Proponents of the reform stress that as long as a defendant perceives a non-zero probability of conviction, a prosecutor can offer a favorable enough plea that the defendant should rationally accept, especially if

^{1}This fact has also been acknowledged by the Supreme Court of the United States. In Santobello v New York, 404 U.S. 257, 260 (1971), the Court wrote, “If every charge were subject to full trial, the states and the federal government would need to multiply by many times the number of judges and court facilities.”

^{2}Gazal-Ayal (2005) advocates the capped plea discounts analyzed in this paper. Guidorizzi (1998) argues for a fixed, proportional discount to be given in exchange for a guilty plea, ideally tailored such that innocent defendants exercise their right to trial and guilty defendants plead guilty. Covey (2007) argues that floors on plea offers are difficult to implement due to charge bargaining and instead advocates trial ceilings to be set based on an initial plea offer. Despite their differences, all of these reforms impose a cap on plea discounts.
risk averse. Hence, if there is no cap on plea discounts, even innocent defendants against whom there is little evidence might plead guilty.

Those opposed to plea bargain reform do not deny that plea bargains might induce pleas from the innocent. Rather, they argue that the fact that the innocent defendants might plead guilty is not inherently troubling, and that preventing innocent defendants from settling would only create further harm. The key idea behind this position is the traditional economic argument that plea bargaining is Pareto efficient due to its voluntary nature (Gazal-Ayal, 2005). Because both prosecutors and defendants are never forced to agree to any specific plea bargain, plea bargaining should occur only when its terms are beneficial for all parties involved. Therefore, according to the traditional economic argument, if capped plea discounts prevent a case from settling, all parties should be left weakly worse off. The argument that plea bargains benefit all parties has been acknowledged by both the Supreme Court of the United States\(^3\) and a wide range of scholars (Scott and Stuntz, 1992; Grossman and Katz, 1983).

Recent scholarship, however, has suggested that the traditional economic argument fails to account for prosecutorial resource constraints and that the welfare implications of plea bargain restrictions may be more nuanced (Gazal-Ayal, 2005; Bar-Gill and Ben-Shahar, 2009; Bar-Gill and Gazal-Ayal, 2006). In a notable legal paper, Gazal-Ayal (2005) argues that capped plea discounts would lead to prosecutors pressing charges against fewer innocent defendants. Gazal-Ayal stresses that because prosecutors only have the resources to try a fraction of their cases, they mostly press charges in cases that can be settled with high probability. Because capped plea discounts prevent prosecutors from securing pleas in weak cases, Gazal-Ayal asserts that capped plea discounts should lead prosecutors to pursue fewer innocent defendants.\(^4\) Bar-Gill and Gazal-Ayal (2006) formalize this argument in a simple model of case selection in which a prosecutor charges

\(^{3}\)See Santobello v New York, 404 U.S. 257, 260 (1971), the Court held, “Disposition of charges after plea discussions is not only an essential part of the process but a highly desirable part for many reasons.”

\(^{4}\)This relies on the intuitive assumption that evidence against innocent defendants tends to be weaker than evidence against guilty defendants.
defendants while subject to a resource constraint.

Like Gazal-Ayal (2005) and Bar-Gill and Gazal-Ayal (2006), I investigate how resource constraints affect the welfare implications of capped plea discounts. However, I do so in novel context, both theoretically and practically. Bar-Gill and Gazal-Ayal (2006) model a situation in which a prosecutor selects cases before bargaining with her defendants. Their model of pre-bargaining activity involves no interaction between the prosecutor and the defendants and therefore is not game-theoretic. My paper, on the other hand, investigates capped plea discounts within the context of game-theoretic bargaining. My paper thus relates to the broad economic literature on bargaining, which I summarize now.

2.2 Related Models

There is extensive literature on the settlement of both civil and criminal litigation (Spier, 2007; Daughety and Reinganum, 2008). Landes (1971) provides the first formal analysis of plea bargaining. Landes assumes that a prosecutor maximizes the sum of the total sentences served by a group of defendants by allocating limited resources among her cases. While Landes’s model is a landmark analysis of plea bargaining, Grossman and Katz (1983) offer the first game-theoretic model of plea bargaining. Grossman and Katz (1983) model plea bargaining using a screening model, which is a canonical model of bilateral bargaining under asymmetric information famously applied to civil litigation by P’ng (1983) and Bebchuk (1984). In a screening model, there is an informed party who knows his type and an uninformed party who knows only the probability distribution of the informed party’s type. The uninformed party makes a take-it-or-leave-it settlement offer to the informed party before trial. The settlement offer acts as a “screen” because the informed party’s response to the settlement may provide information about his type.5

In Grossman and Katz (1983), an uninformed prosecutor offers a take-it-or-leave-it

5 The screening model can be contrasted to the signaling model, in which the informed party makes an offer to the uninformed party.
plea to the defendant, who knows whether he is guilty. Guilty types are more likely to be convicted at trial than innocent types, so the defendant may be more likely to accept a plea offer when he is guilty than innocent (Grossman and Katz, 1983). As a result, the prosecutor may gain information about the defendant’s type from his response to the plea. Unlike in Landes (1971), the prosecutor is socially interested, meaning that she wants to punish guilty but not innocent types. Grossman and Katz’s screening framework has strongly influenced subsequent models of plea bargains, and several prominent models of plea bargaining, including those in Reinganum (1988) and Baker and Mezzetti (2001), have built upon Grossman and Katz’s model. Even models that deviate from the screening framework generally maintain the bilateral, asymmetric structure featured in Grossman and Katz (1983) (Reinganum, 2000).

Bar-Gill and Ben-Shahar (2009), however, abandon both the bilateral structure and the asymmetric information of traditional plea bargaining models. Bar-Gill and Ben-Shahar consider a multilateral game in which a prosecutor offers $N$ take-it-or-leave-it pleas to $N$ defendants. Information is symmetric, as the prosecutor and the defendants agree on the expected sentences at trial and pleas offers are publicly observable. The unique feature of this model is that the prosecutor has limited trial capacity: although she bargains with multiple defendants, she has the resources to try only one of them. Although earlier models of plea bargaining, such as Reinganum (1988) and Baker and Mezzetti (2001), assume that the prosecutor has finite resources, Bar-Gill and Ben-Shahar (2009) are the first to explicitly limit the number of cases a prosecutor can try.

The trial constraint of Bar-Gill and Ben-Shahar (2009) cleverly connects plea bargaining to the contracting with externalities literature. The contracting with externalities literature studies multilateral games in which the agents’ utilities are functions not only of their own actions but also of the actions of other agents (Segal, 1999). In Bar-Gill and Ben-Shahar (2009), defendants contract with externalities because the decision of one defendant to plead guilty shifts the trial threat onto other defendants. Bar-Gill and
Ben-Shahar show that the prosecutor can exploit these negative externalities to induce all \( N \) defendants to settle for their expected trial sentence despite the fact that she can only try one defendant.\(^6\) This result follows from the fact that if the prosecutor makes her trial priority list clear, the prosecutor can induce all \( N \) defendants to iteratively reason that they will be tried if they reject their pleas.\(^7\) Bar-Gill and Ben-Shahar (2009) illustrate that when prosecutors are resource constrained, plea bargaining may be in each defendant’s individual interest but against the defendants’ collective interest.

My model novelly combines the asymmetric screening framework of Grossman and Katz (1983) and the multilateral framework of Bar-Gill and Ben-Shahar (2009). While the combination of asymmetric information and bargaining externalities complicates equilibrium analysis, both elements are essential aspects of plea bargaining. Like Bar-Gill and Ben-Shahar (2009), I assume that the prosecutor maximizes the sum of the defendants’ sentences,\(^8\) which aligns my paper with most models of civil litigation.\(^9\) Additionally, because I assume that trial is not always profitable for the prosecutor, my model relates closely with the Nalebuff’s model of civil litigation (Nalebuff, 1987). Nalebuff (1987) shows that as a settlement offer becomes more lenient, the types of defendants who reject the plea become less attractive for the plaintiff to try. Thus, there may exist a minimum settlement offer for which the plaintiff can credibly threaten to trial with probability 1 upon the rejection of the plea in equilibrium. This credibility constraint leads to the somewhat counterintuitive result that offering a more lenient settlement may not

\(^6\)The result that the principal can exploit negative externalities between agents is a common result in the contracting with externalities literature. For example, Rasmusen et al. (1991) and Segal and Whinston (2000) famously show that a monopolist may be able to profitably exclude competition through exclusive contracts with buyers. Segal and Whinston demonstrate that if the monopolist can secure a sufficient number of contracts such that it is no longer profitable for the competitor to end the market, then the monopolist can discriminate and offer “bad” contracts to the rest of agents. The fact that a critical mass of agents will sign the contract imposes a negative externality on the rest of the agents, allowing the monopolist to profit through discrimination.

\(^7\)For example, the defendant who is the prosecutor’s first priority will plead guilty, knowing he will be tried. Knowing that the first defendant will plead guilty, the second priority knows he will also be tried. The process iterates.

\(^8\)I justify my choice of the prosecutor’s objective function when introducing the model in Section 3.

\(^9\)See Spier (2007) for a broad review of models of civil litigation.
necessarily induce more defendants to accept the offer in equilibrium.

Finally, Section 6 directly relates to the literature on externalities and sequential offers (Che and Spier, 2008; Hua and Spier, 2005; Daughety and Reinganum, 2002; Che and Yi, 1993). In particular, Che and Spier (2008) analyze a model of multilateral civil litigation under both simultaneous and sequential offers.\footnote{Che and Spier (2008) analyze a model in which multiple plaintiffs sue the same defendant and share the costs of litigation. They assume symmetric information for their main models and then consider asymmetric information in an extension.} Che and Spier show that the principal may be able to better exploit externalities when offers are sequential, which I replicate in Section 6.
3 The Model

A risk-neutral prosecutor offers a pair of take-it-or-leave-it plea offers to two risk-neutral defendants, $D_1$ and $D_2$, before costly trial. The defendants have two possible types: guilty and innocent. I denote the type of $D_i$ as $\tau_i \in \{g, i\}$, where $g$ denotes guilt and $i$ denotes innocence. The types of the two defendants are independent and identically distributed such that $P(\tau_1 = g) = P(\tau_2 = g) = \rho$. Each defendant observes his own type before bargaining begins but does not observe the type of the other defendant. The prosecutor does not observe the type of either defendant. All parties know the joint probability distribution of the defendants’ types.

The timing of the model is as follows. First, the prosecutor makes a pair of take-it-or-leave-it plea offers to the defendants. I denote the pair of plea offers as $s = \{s_1, s_2\}$, where $s_i$ is the plea offer to $D_i$. Pleas are publicly observable, meaning that defendants observe not only their own plea offer but also the offer to the other defendant. Second, the defendants simultaneously respond to their pleas. If $D_i$ accepts his plea offer, he serves a sentence of $s_i$. If he rejects, he is eligible for trial. Third, the prosecutor decides whether to proceed to trial against the defendants who rejected their pleas. The prosecutor’s trial capacity is the crux of the model. I begin by supposing that the prosecutor has the resources to try both defendants (Section 4) and then suppose that the prosecutor only has the resources to try one of the two defendants (Section 5). Fourth is the trial stage. As is typical in the litigation literature, trial incurs effort costs for both parties. I denote the trial costs for the prosecutor and the defendants as $c_p$ and $c_d$, respectively.\textsuperscript{11} Defendants are more likely to be convicted at trial if guilty than if innocent, and both defendants have the same expected trial sentences conditional on type. Thus, if tried, the expected sentence of a defendant is $X$ if he is guilty and $\pi X$ if he is innocent, where

\textsuperscript{11}There is qualitative evidence that these efforts costs can be quite high. Alschuler (1983) reports that prosecutors constantly worry about how long trials will last and are much more eager to settle cases that would require lengthy trials. Bebchuk (1984) reports that defendants may often be held in jail for long periods of time before trial. Thus, waiting for trial incurs not only additional lawyering costs but also may carry an implicit increase in sentencing.
\[ \pi \in (0, 1). \] Note that the game may end before the third or fourth stage if both defendants accept their pleas or if the prosecutor drops her cases against the defendants.

The prosecutor maximizes the sum of the defendants’ sentences, trial costs held equal. The objective functions of prosecutors vary throughout the plea bargaining literature. Grossman and Katz (1983) and related papers on plea bargaining suppose that the prosecutor is socially interested, maximizing the sentences of guilty types while minimizing the sentences of innocent types (Spier, 2007). Landes (1971) and Bar-Gill and Ben-Shahar (2009), on the other hand, suppose that prosecutors simply maximize the sum of the defendants’ sentences. While the empirical literature on prosecutorial incentives is somewhat sparse, several studies support my choice of objective function. In a longitudinal study of 570 chief federal prosecutors, Boyland (2005) finds that prosecutors who secure longer sentences tend to have more successful careers. Additionally, in an analysis of how prosecutors allocate drug crimes between state and federal systems, Glaeser et al. (2000) find evidence that prosecutors maximize their private human capital. There is also qualitative evidence that determining the guilt of defendants is a relatively small concern for prosecutors (Alschuler, 1983). In an extensive series of interviews with prosecutors, Alschuler (1983) finds that most prosecutors deem the “quasi-judicial” role of assessing the guilt of a defendant to be a secondary concern. To be clear, I do not assert that prosecutors are apathetic about the guilt of defendants. Rather, given that the empirical literature indicates that prosecutors often prioritize their careers and that securing large sentences improves career prospects, I merely suggest that modeling the prosecutor’s objective function as strictly increasing in the length of sentences captures a primary, career-based incentive that prosecutors have.

Less controversially, defendants minimize the sum of their individual trial costs and sentences at trial. I assume that all parties are risk neutral. Letting \( \sigma_i \) denote sentence served by \( D_i \) and \( I_i^T \) be an indicator variable for whether \( D_i \) is tried, the respective ex
post utilities of the prosecutor and $D_i$ can be expressed

$$U_p^{ex\ post} = \sum_{i=1}^{2} (\sigma_i - c_p I_i^T)$$

(1)

$$U_{d_i}^{ex\ post} = -\sigma_i - c_d I_i^T.$$  

(2)

Like Nalebuff (1987), I assume that trial has a positive expected value for the prosecutor based on her prior beliefs\(^{12}\) about the defendants’ types but does not have a positive expected value conditional on the defendant being innocent. Formally, I make the following assumption.

**Assumption 3.1.** $c_p \in (\pi X, \rho X + (1 - \rho)\pi X)$.

The lower bound, $\pi X$, ensures that the cost of trial for the prosecutor is greater than the expected benefit of trying an innocent defendant. This assumption is similar to that made in Nalebuff (1987) and also reflects the fact the Brady Rule legally obligates prosecutors to release exculpatory evidence, essentially forcing them to drop cases against defendants known to be innocent (Legal Information Institute, 2017). The upper bound of $\rho X + (1 - \rho)\pi X$ ensures that trial has a positive expected value based on the prosecutor’s prior beliefs. When this upper bound is violated, the model becomes degenerate.\(^{13}\)

Finally, I define my social welfare criterion.

**Definition 3.1 (Social objective function).** I define the social objective function as

$$U_s = \lambda U_i - U_y - \epsilon T$$

(3)

\(^{12}\)To be clear, her prior beliefs about the defendants’ types are that each defendant is guilty with probability $\rho$ and that types are independent. I use prior beliefs to refer to the prosecutor’s knowledge of the joint probability distribution of the defendants’ types.

\(^{13}\)This model becomes degenerate because defendants are always weakly more likely to accept a plea when guilty than innocent. So, after observing a rejection, the prosecutor’s updated belief about a defendant’s probability of guilt must be lower than her prior belief. Therefore, if the prosecutor does not try under her prior, she will never try, leading to degeneracy.
where \( U_i \) is the expected utility of an innocent defendant in equilibrium, \( U_g \) is the expected utility of a guilty defendant in equilibrium, \( T \) is the expected number of trials in equilibrium, and \( \lambda \) and \( \epsilon \) are positive constants, with \( \epsilon \) being arbitrarily small.

Note that the expected utility of “a defendant” is simply the average of \( D_1 \)’s and \( D_2 \)’s expected utilities. This social utility function reflects the fact that society benefits when the guilty defendants are punished but suffers when innocent defendants are punished.\(^{14}\) This social preference can be justified by both the wish to deter crime and by the societal desire to give people their just deserts. Similar social utility functions are common throughout the plea bargaining literature (Baker and Mezzetti, 2001; Grossman and Katz, 1983; Reinganum, 1988). My social utility function also implies that society prefers to save resources when possible but that this desire is secondary to its desire to punish the guilty and avoid punishing the innocent. While this assumption clearly does not hold when resource expenditure is extreme, I suggest that this priority should hold when the number of trials is manageable for the prosecutor and the courts. Thus, my social utility function should at least be locally true.

The prosecutor’s utility is not directly modeled in the social utility function because her utility is zero-sum with the utility of the defendants, and is thus in direct conflict with the desire not to punish the innocent. Moreover, I assert that the role of the criminal justice system is to properly adjudicate between the guilty and the innocent and that the career concerns of prosecutors should be a minor secondary concern.

\(^{14}\)One may note that society only wants to punish guilty defendants only to a certain extent; the social objective function assumes that \( X + c_d \), which is the maximum expected sentence the defendant can serve is not excessive punishment.
4 Analysis Without Resource Constraint

I now analyze the game when the prosecutor has the resources to try both of the defendants. This section serves as a reference before introducing the prosecutorial resource constraint in Section 5.

I find that capped plea discounts prevent innocent defendants from pleading guilty in equilibrium, as proponents of the reform suggest. However, the restriction leaves all parties weakly worse off in equilibrium, as the traditional economic argument predicts. This result holds regardless of whether the prosecutor can commit \textit{ex ante} to a trial strategy.

The following proposition helps simplify the analysis of the game.

**Proposition 4.1 (Representation as two bilateral games).** \textit{When the prosecutor can try both of her cases, the multilateral game is equivalent to two independent bilateral games.}

\textit{Proof.} This proposition quickly follows from two points. First, the defendants’ types are independent, so the prosecutor’s beliefs about one defendant’s guilt are independent of the actions of the other defendant. Second, the prosecutor is not trial constrained. Thus, there are no bargaining externalities, as one defendant’s plea acceptance does not affect whether the prosecutor can try the other defendant.

Because of this equivalence, I analyze the bilateral game without loss of generality. For an extensive form representation of the bilateral game, see Figure 1. I use the notation $\tau \in \{g, i\}$ to denote the type of the defendant in the bilateral model and $s$ to denote the plea offer, dropping the subscripts necessary in the multilateral model. To be clear, I use $U_p$ throughout this section to refer to the prosecutor’s payout from the bilateral game, not the multilateral game. This bilateral model is a classic bilateral screening model, and thus serves primarily as a benchmark for Section 5 rather than as a novel theoretical contribution.
4.1 Perfect Information Benchmark

I begin by analyzing the game under perfect information as a benchmark. To be clear, types are publicly known under perfect information. Replicating the standard result in the law and economics literature on bilateral bargaining with perfect information, the prosecutor and the defendant settle and avoid costly trial with probability 1 in equilibrium.

**Proposition 4.2.** Under perfect information,

(i) in subgame perfect equilibrium, the prosecutor tries the defendant if he is guilty but not if he is innocent. The defendant accepts if guilty only if \( s \leq X + c_d \) and rejects all \( s \) when innocent. The prosecutor thus offers \( s = X + c_d \) to guilty types and any plea to innocent types.
(ii) equilibrium is the socially optimal outcome of the game.

Proof. Under perfect information, the prosecutor does not try innocent types by Assumption 3.1. Knowing this, innocent defendants reject all $s > 0$, and the prosecutor is indifferent among pleas. Guilty types know that they will be tried, and thus accept pleas less than their expected disutility at trial. The prosecutor offers the maximum plea the guilty defendant will accept, which is $s = X + c_d$.\(^{15}\)

This equilibrium is socially optimal because innocent defendants are not punished, guilty defendants are maximally punished, and trial never occurs. Guilty defendants are maximally punished because they can always opt for trial, meaning that their worst case expected utility is $-X - c_d$.

The equilibrium payouts are

$$U_p = \rho(X + c_d), \quad U_g = -X - c_d, \quad U_i = 0, \quad (4)$$

and maximum social utility achievable is

$$U_s = X + c_d. \quad (5)$$

Because equilibrium is socially optimal under perfect information, reform is clearly unnecessary in this situation.

If the prosecutor is able to commit *ex ante* to a trial strategy, she is able to credibly threaten trial against innocent types. When *ex ante* commitment is permitted, the prosecutor induces pleas from both innocent and guilty types, shifting the game away from its social optimum. Instituting capped plea discounts by restricting $s > \pi X + c_d$ returns equilibrium to its socially optimum.

\(^{15}\)I assume that the defendant accepts at his indifferent point. This assumption does not change the results, as the prosecutor can offer $s = \pi X + c_d - \epsilon$ for arbitrarily small $\epsilon$ instead.
Proposition 4.3. Under perfect information, if the prosecutor may commit ex ante to a trial strategy,

(i) she commits to trial in equilibrium. Then, she offers $X + c_d$ to guilty types and $\pi X + c_d$ to innocent types, both of which are always accepted.

(ii) capping plea discounts by restricting $s > \pi X + c_d$ returns equilibrium to its social optimum.

Proof. If the prosecutor commits to trial, the defendant will accept any plea less than or equal to his expected disutility at trial. It is clearly optimal for the prosecutor to extract both the expected sentence and the cost of trial from both types of defendants without suffering the cost of trial herself. Because innocent defendants have disutility $\pi X + c_d$, equilibrium has shifted away from its social optimum.

By restricting $s > \pi X + c_d$, one prevents the prosecutor from inducing pleas from innocent types. The prosecutor then drops cases against the innocent defendant, as committing to trial cannot induce pleas and trial itself have negative expected value. This returns the game the social optimal equilibrium described in Proposition 4.1. □

This benchmark illustrates that capped plea discounts improve the welfare of innocent defendants under perfect information only if the prosecutor can threaten trial against innocent defendants.

4.2 Characterization of Sequential Equilibria

I now return to the model of asymmetric information and characterize the sequential equilibria\textsuperscript{16} of the asymmetric game using backwards induction. For definitions of the game theory concepts used, see Appendix C.

\textsuperscript{16}To be clear, sequential equilibrium refers to a common equilibrium refinement in dynamic, asymmetric games. I define sequential equilibrium in Appendix C. Sequential equilibrium does not imply that offers are made sequentially (for analysis of the game when offers are made sequentially, see Section 6).
4.2.1 The Prosecutor’s Response

Suppose that the defendant has rejected $s$. Although trial has positive expected value for the prosecutor based on her prior beliefs, trial may not have a positive expected value conditional on the rejection of the plea.

Let $\hat{p}^*_p(r|\tau)$ denote the prosecutor’s belief about the probability that the defendant rejects $s$ if he is type $\tau$. The prosecutor tries the defendant only if

$$c_p \leq \frac{\pi \hat{p}^*_p(r|i)(1 - \rho) + \hat{p}^*_p(r|g)\rho}{\hat{p}^*_p(r|i)(1 - \rho) + \hat{p}^*_p(r|g)\rho} X. \tag{6}$$

The right side of inequality (6) is the prosecutor’s posterior belief about the expected sentence at trial given that $s$ has been rejected. Note that this condition can be rewritten using Bayes’ Rule as

$$\hat{p}^*_p(g|r) \geq \frac{c_p/X - \pi}{1 - \pi}. \tag{7}$$

Inequality (7) simply states that the prosecutor tries the defendant only if he is sufficiently likely to be guilty after rejecting his plea.

When equality holds, the prosecutor is indifferent about proceeding to trial. If the defendant rejects with probability 1 when innocent, the prosecutor is indifferent about going to trial in equilibrium if and only if the defendant accepts with probability $\rho + \pi(1 - \rho) - c_p/X \rho(1 - c_p/X)$ when guilty. I let

$$q_0 = \frac{\rho + \pi(1 - \rho) - c_p/X}{\rho(1 - c_p/X)}. \tag{8}$$

4.2.2 The Defendant’s Response

Suppose that the prosecutor has offered a plea $s$ to the defendant. Let $\hat{p}^*_d(t|r)$ be the defendant’s posterior belief about the probability of trial after he has rejected $s$. The

\[\text{Using similar notation as above, I let } \hat{p}^*_d(\tau|r) \text{ be the prosecutor’s posterior belief about the probability that the defendant is type } \tau \text{ conditional upon him rejecting } s.\]
defendant accepts when innocent only if

\[ s \leq (\pi X + c_d)\hat{p}_d^s(t|r). \] (9)

and accepts when guilty only if

\[ s \leq (X + cd)\hat{p}_d^s(t|r). \] (10)

The right sides of the inequalities (9) and (10) are the expected sentence the defendant believes he faces conditional on the rejection of \( s \).

**Lemma 4.1.** (Subgame equilibria) Fix

(i) \( s \in (0, \pi X + c_d] \). Then there exists a subgame equilibrium in which the defendant accepts \( s \) regardless of type.

(ii) \( s \in (0, X + c_d] \). Then there exists a subgame equilibrium in which the defendant rejects if innocent and mixes\(^{18}\) if guilty, accepting with probability \( q_0 \). If the plea is rejected, the prosecutor tries with probability \( \frac{s}{X + c_d} \).

(iii) \( s \geq X + c_d \). Then there exists a subgame equilibrium in which the defendant rejects \( s \) and the prosecutor tries upon rejection.

No other subgame equilibria exist for positive \( s \neq X + c_d \).

**Proof.** See Appendix A.

This lemma has two main implications. First, if a plea offer is less than or equal to the innocent defendant’s expected disutility of trial, \( \pi X + c_d \), the defendant may accept the plea regardless of type. This pooling subgame equilibrium occurs only if the defendant’s off-path belief about trial is sufficiently high. Second, if \( s \in (\pi X + c_d, X + c_d) \),

\(^{18}\)The defendant “mixing” refers to the defendant playing a mixed strategy, in which he both accepts and rejects with positive probability.
meaning that the plea is greater than the expected disutility of trial for innocent types but less than that for guilty types, there exists a unique subgame equilibrium in which the defendant rejects when innocent and mixes when guilty. The innocent defendant must reject because the plea is higher than his expected disutility at trial. Given that the innocent type rejects, the guilty type must accept with probability $q_0$, which is the probability that makes the prosecutor indifferent about pursuing trial. If the guilty type were to accept more frequently, the prosecutor would never try, allowing all types to deviate by rejecting. If the guilty type were to accept less frequently, the prosecutor would always try, encouraging the guilty type to deviate by accepting any $s < X + c_d$. This result implies that the prosecutor may not be able to induce a higher rate of acceptance by lowering her plea offer once subgames are in equilibrium, replicating the finding of Nalebuff (1987).

4.2.3 The Prosecutor’s Offer

Let $s^*_A$ denote the maximum $s$ for which the defendant accepts regardless of type in subgame equilibrium. By Lemma 4.1, $s^*_A \in (0, \pi X + c_d] \cup \emptyset$.

**Lemma 4.2.** (Sequential equilibrium) In the bilateral game, the potential on-path strategies played in sequential equilibrium are as follows:

(i) Pooling Equilibrium: If $s^*_A \geq \rho q_0(X + c_d)$, the prosecutor offers $s^*_A$ in equilibrium, which the defendant accepts regardless of type.

(ii) Semi-Separating Equilibrium: If $s^*_A \leq \rho q_0(X + c_d)$, the prosecutor offers $s = X + c_d$, and tries with probability 1 upon rejection. The defendant rejects when innocent and accepts with probability $q_0$ when guilty.

*Proof.* See Appendix A.

Lemma 4.2 states that prosecutor pursues one of two strategy types in sequential equilibrium. When the percentage of innocent defendants and prosecutor’s cost of trial
are high,\textsuperscript{19} the prosecutor offers the maximum plea that will always be accepted by the defendant. On the other hand, when these parameters are low, the prosecutor offers the maximum plea that guilty defendants will accept with positive probability.

Note that innocent defendants plead guilty in the pooling equilibrium but not in the semi-separating equilibrium. Thus, as the defendant pool becomes more heavily innocent, the more likely it is that innocent defendants will plead guilty in equilibrium.

\subsection*{4.2.4 Welfare in Equilibrium}

In the pooling equilibrium, the equilibrium payouts are

\begin{equation}
U_p = s_A^*, \quad U_g = -s_A^*, \quad U_i = -s_A^*
\end{equation}

and social utility is

\begin{equation}
U_s = (1 - \lambda)s_A^*.
\end{equation}

Relative to the socially optimum outline in Proposition 4.2, guilty defendants are strictly better off while innocent defendants are strictly worse off. The prosecutor may or may not be better off in this new equilibrium, depending on relation between $s_A^*$ and $\rho(X + c_d)$.\textsuperscript{20}

In the semi-separating equilibrium, the equilibrium payouts are

\begin{equation}
U_p = \rho q_0(X + c_d), \quad U_g = -X - c_d, \quad U_i = -\pi X - c_d
\end{equation}

\textsuperscript{19}Note that $q_0$ decreases with $c_p$.

\textsuperscript{20}An interesting implication of this result is that the prosecutor may prefer ignorance to knowledge about the defendant’s type. If $s_A^* > \rho(X + c_d)$, then it is better for the prosecutor not to know the defendant’s type so that she can credibly threaten trial and extract pleas from innocent defendants. Because $s_A^* \leq \pi X + c_d$, this can only occur when

\begin{equation}
\pi X + c_d > \rho(X + c_d),
\end{equation}

or when $\pi$ large relative to $\rho$, and $c_d$ is large relative to $X$.  

21
and social utility is

\[ U_s = -\lambda (\pi X + c_d) + (X + c_d) - \epsilon (1 - \rho q_0). \]  \hspace{1cm} (15)

Relative to the socially optimal equilibrium, the semi-separating equilibrium does not change the welfare of guilty defendants but leaves innocent defendants and the prosecutor strictly worse off.

### 4.3 The Effect of Capped Plea Discounts

When the pooling equilibrium occurs, innocent defendants plead guilty with probability 1. Thus, those concerned about the innocence problem might advocate capping the plea discount to prevent the pooling equilibrium. Restrict \( s > \pi X + c_d \) is the most intuitive way to cap the plea discount, as it ensures that innocent defendants will never plead guilty.

**Proposition 4.4.** When one restricts \( s > \pi X + c_d \), the semi-separating equilibrium is the unique sequential equilibrium of the bilateral game. This restriction prevents innocent defendants from pleading guilty in equilibrium but leaves all parties weakly worse off.

**Proof.** By Lemma 4.1, if one restricts \( s > \pi X + c_d \), the prosecutor cannot offer a plea that will be accepted by the defendant regardless of type. Thus, \( s^*_A = \emptyset \). By Lemma 4.2, it follows that the screening equilibrium is the unique sequential equilibrium after the restriction.

By Lemma 4.2, the restriction binds only if \( s^*_A \geq \rho q_0 (X + c_d) \). The changes in each party’s utility after the restriction are

\[ \Delta U_p = \rho q_0 (X + c_d) - s^*_A \]  \hspace{1cm} (16)

\[ \Delta U_g = -(X + c_d) + s^*_A \]  \hspace{1cm} (17)
\[ \Delta U_i = -(\pi X + c_d) + s_A^* \] (18)

which are all non-positive.

Proposition 4.4 supports the traditional economic argument that plea bargains are efficient contracts and that preventing them leaves all parties worse off. While capped plea discounts may shift equilibrium so that the defendant no longer pleads guilty in equilibrium, it does not help innocent defendants because they are tried upon the rejection of their plea. The directional change in social utility is unclear, as

\[ \Delta U_s = -\lambda(\pi X + c_d) + (X + c_d) - (1 - \lambda)s_A^* - \epsilon(1 - (\rho q_0)^2) \] (19)

has an ambiguous sign that depends on the magnitude of \( \lambda \). However, this benchmark illustrates that when prosecutors are not resource constrained, those concerned primarily with the welfare of innocent defendants should not support capped plea discounts.

Notably, however, the partial ban of capped plea discounts is strictly preferable to a complete ban on plea bargains, as the two reforms lead to the same outcomes for the defendants and the full ban requires more frequent trial.

**Corollary 4.4.1.** When the prosecutor is not resource constrained, capped plea discounts achieve the same welfare outcomes for defendants as a total ban on plea bargaining. However, capped plea discounts lead to fewer trials than a complete ban. Thus, capping plea discounts is strictly preferable to a total ban.

This follows from the fact that when pleas are banned, the prosecutor tries with probability 1. This yields welfare outcomes

\[ U_p = \rho X + (1 - \rho)\pi X - c_p, \quad U_y = -X - c_d, \quad U_i = -\pi X - c_d \] (20)

and social welfare

\[ U_s = -\lambda(\pi X + c_d) + (X + c_d) - \epsilon, \] (21)
which is strictly worse than the outcome under capped plea discounts.

This benchmark illustrates that when a prosecutor can try all of her cases, capping plea discounts only benefits society if the gain from further punishing the guilty outweighs the potential harm to the innocent. However, the reform achieves better welfare outcomes than abolition of plea bargains.

4.4 Allowing Ex-Ante Commitment

Finally, I consider the game when the prosecutor may commit *ex ante* to trial. I show that capped plea discounts leave guilty defendants and the prosecutor weakly worse off but does not affect the welfare of innocent defendants.

**Proposition 4.5.** If the prosecutor may commit *ex ante* to trial, equilibrium of the bilateral asymmetric game is as follows.

(i) If \( \pi X + c_d \geq \rho (X + c_d) + (1 - \rho)(\pi X - c_p) \), the prosecutor commits to trial *ex ante* and offers \( s = \pi X + c_d \). The defendant accepts with probability 1 regardless of type.

(ii) If \( \pi X + c_d \leq \rho (X + c_d) + (1 - \rho)(\pi X - c_p) \), the prosecutor commits to trial *ex ante* and offers \( s = X + c_d \). The defendant accepts when guilty and rejects when innocent.

**Proof.** Once the prosecutor has committed to trial, the defendant accepts any offer less than or equal to his expected disutility at trial.\(^{21}\) The prosecutor’s two reasonable strategies are to offer the expected disutility of the innocent or guilty types, as described in the equilibria above.

The prosecutor is weakly better off committing. By committing and offering \( \pi X + c_d \), the prosecutor does weakly better than she does in the pooling equilibrium, achieving \( U_p = \pi X + c_d \geq s^*_A \). By committing and offering \( X + c_d \), the prosecutor does strictly better

\(^{21}\)I again assume that the defendant accepts at his indifference point because the prosecutor can offer a plea arbitrarily close to the indifference point that will be accepted.
than in the semi-separating equilibrium, achieving \( U_p = \rho(X + c_d) + (1 - \rho)(\pi X - c_p) \).

This commitment outcome is strictly better than the semi-separating equilibrium for the prosecutor because, in both, the prosecutor tries the defendant when innocent with probability 1. With commitment, however, guilty types do not mix, and the prosecutor extracts sentence \( X + c_d \) from guilty types without ever going to trial.

Note that fully committing dominates committing to try with probability \( 1 - \gamma \). When the prosecutor commits to trial with probability \( 1 - \gamma \), her the utilities from her best strategies are scaled by a factor of \( 1 - \gamma \). This follows from using the same reasoning that was used determine to prosecutor’s utility when committing fully.

Once again, it is possible that the innocence problem occurs in equilibrium, leading to the possibility of binding capped plea discounts.

**Corollary 4.5.1.** When the prosecutor can commit ex ante to trial, capping plea discounts by restricting \( s > \pi X + c_d \) leaves guilty defendants weakly worse off and does not affect innocent defendants, thus increasing social utility.

This corollary follows from the fact that restriction \( s > \pi X + c_d \) forces the prosecutor to offer \( X + c_d \) instead of \( \pi X + c_d \). This leaves guilty defendants strictly worse off but does not affect innocent defendants, as seen in Table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>( U_g )</th>
<th>( U_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commitment, ( s = X + c_d )</td>
<td>(-X - c_d)</td>
<td>(-\pi X - c_d)</td>
</tr>
<tr>
<td>Commitment, ( s = \pi X + c_d )</td>
<td>(-\pi X - c_d)</td>
<td>(-\pi X - c_d)</td>
</tr>
</tbody>
</table>

Table 4.1: Defendant welfare in commitment outcomes

Thus, when prosecutors are able to or are forced to commit to trial, there is some evidence that capping plea discounts is good social policy. However, this potential social benefit comes from losses from guilty defendants, not from improvement for innocent types.
as reformers hope. Hence, plea discounts do not help innocent defendants when the
prosecutor can try both defendants, regardless of whether commitment is feasible.
5 Analysis With Resource Constraint

I now suppose that the prosecutor has the resources to try at most one of the two defendants. The key implication of this resource constraint is that it creates a negative bargaining externality between the defendants. When one defendant is expected to accept his plea, the prosecutor may focus her trial threat on the other defendant. As when the prosecutor can try both defendants, I find that innocent defendants may plead guilty in equilibrium. However, the welfare implications of capped plea discounts for innocent defendants are reversed, as the reform may leave innocent defendants better off in equilibrium. Additionally, even when innocent defendants are left better off, guilty defendants may be left worse off. I show that these results hold regardless of whether the prosecutor can commit ex ante to a trial strategy.

Due to the combination of asymmetric information and the contracting externalities, characterizing sequential equilibrium without refinement is unwieldy and confuses the intuition of the model. Thus, I assume that subgames are prosecutor optimal throughout my analysis.

Assumption 5.1 (Prosecutor optimal subgame refinement). A subgame equilibrium is prosecutor optimal if and only if it is the subgame equilibrium that maximizes the expected utility of the prosecutor. I assume that subgame equilibria are prosecutor optimal throughout the analysis of Section 5.

Because I use a subgame refinement, the results of this section do not reflect all possible equilibria outcomes, as the results of Section 4 do. However, there are two main reasons to believe that subgames may be close to prosecutor optimal. First, it is often noted in the legal literature that prosecutors have significantly more bargaining power than defendants, suggesting that bargaining subgames may favor the prosecutor (Alschuler, 1968; White, 1971). Second, I show later that the prosecutor optimal sub-
game equilibria are those that maximize the chance that defendants plead guilty.\textsuperscript{22} The fact that over 90 percent of convictions are obtained via pleas thus supports the selection of these prosecutor optimal subgames (U. S. Sentencing Commission, 2010; U. S. Department of Justice, 2006).

5.1 Perfect Information Benchmark

I begin by analyzing the game under perfect information as a benchmark. As when the prosecutor is not resource constrained, there exists a socially optimal equilibrium under perfect information in which guilty types are fully punished and innocent defendants are not punished. Notably, this analysis under perfect information mirrors that of Bar-Gill and Ben-Shahar (2009).

**Proposition 5.1.** (i) Under perfect information, there exists a subgame perfect equilibrium in which the prosecutor offers \( s = X + c_d \) to guilty defendants and any plea to innocent defendants. Guilty defendants accept while innocent defendants reject. The prosecutor drops cases against innocent defendants upon rejection.

(ii) The equilibrium described in (i) is the socially optimal equilibrium of the game.

*Proof.* See Appendix A.

This socially optimal equilibrium is attainable despite the resource constraint because guilty defendants may believe that the other defendant always accepts his plea in equilibrium when guilty. Thus, because guilty defendants suffer from a coordination problem, the prosecutor’s resource constraint does not inhibit her ability to threaten trial.

Note that although this is not the unique equilibrium of the game, the prosecutor can ensure this coordination failure if she makes her trial priority list clear. For example, if she makes it publicly known that she will always try \( D_1 \) if he rejects when guilty, then she can induce a plea from \( D_1 \). Knowing that \( D_1 \) will plead out if guilty, \( D_2 \) also accepts his

\textsuperscript{22}See Corollary 5.2.1.
plea when guilty. Note that this commitment is credible, as the prosecutor is indifferent between trying $D_1$ and $D_2$ when both defendants are guilty. This finding replicates the result of Ben-Shahar and Bar-Gill (2009).

As in Section 4, when the prosecutor can commit *ex ante* to trial, she can induce pleas from innocent types. Restricting $s > \pi X + c_d$ returns equilibrium to its social optimum.

**Proposition 5.2.** *Under perfect information, if the prosecutor may commit *ex ante* to a trial strategy,*

(i) she commits to trying $D_1$ if he rejects his plea and to trying $D_2$ if $D_1$ accepts his plea.

The prosecutor then offers innocent defendants $s_i = \pi X + c_d$ and guilty defendant $s_i = X + c_d$, and both defendants accept regardless of their types.

(ii) instituting a cap on the plea discount by restricting $s > \pi X + c_d$ returns equilibrium to its social optimum.

*Proof.* $D_1$ knows that he will be tried, so he accepts any plea less than or equal to his expected disutility at trial. $D_2$ knows that $D_1$ will plead guilty, so he accepts any plea less than or equal to his expected disutility at trial. The prosecutor clearly cannot do better, as she extracts the full expected disutility of both defendants without having to go to trial.

By restricting $s > \pi X + c_d$, the prosecutor can no longer induce pleas from innocent types, even with commitment. This returns equilibrium to its social optimum.

By committing to trial, the prosecutor can extract the expected trial disutility of both defendants, regardless of type. The prosecutor is not limited by her resource constraint because she can cause all defendants to iteratively reason that they will be tried. Thus, in this perfect information benchmark with commitment, capped plea discounts improve social welfare by preventing the prosecutor from inducing the innocent defendant to plead guilty, returning equilibrium to its social optimal state in which only guilty defendants are punished.

29
5.2 Characterization of Sequential Equilibria

I now characterize the sequential equilibria of the asymmetric game using backwards induction. Throughout, I use subscripts to denote events for specific defendants. For example, $g_i$ denotes the event of $D_i$ being guilty, and $r_i$ denotes the event of $D_i$ rejecting his plea. I denote the beliefs of party $i$ given fixed $\bar{s}$ as $\hat{p}_i^\bar{s}(.)$.

5.2.1 The Prosecutor’s Response

Suppose that $D_1$ has rejected $s_1$ but that $D_2$ has accepted. Because types are independent, the prosecutor’s decision to try $D_1$ is parallel to her decision in the bilateral game. She tries $D_1$ only if

$$c_p \leq \frac{\pi \hat{p}_p^\bar{s}(g_1|r_1)(1 - \rho) + \hat{p}_p^\bar{s}(r_1|g_1)\rho X}{\hat{p}_p^\bar{s}(r_1|i_1)(1 - \rho) + \hat{p}_p^\bar{s}(r_1|g_1)\rho X}. \tag{22}$$

The right side of inequality (22) is the prosecutor’s posterior belief about the expected value of trying $D_1$. Note that the prosecutor’s beliefs about $D_1$ are functions not only of $s_1$ but also of $s_2$.

The prosecutor’s decision is slightly more complicated if both defendants reject their pleas. If both defendants reject their pleas, the prosecutor tries $D_i$ only if

$$\hat{p}_p^\bar{s}(g_i|r_i) \geq \max\{\hat{p}_p^\bar{s}(g_j|r_j), \frac{c_p/X - \pi}{1 - \pi}\}. \tag{23}$$

This necessary condition reflects that the prosecutor tries the defendant who is more likely to be guilty conditional on the rejection, provided that trying that defendant has positive expected value. As before, the prosecutor is indifferent about trying a defendant if he rejects when innocent and mixes when guilty, accepting with probability $\frac{\rho + \pi(1 - \rho) - c_p/X}{\rho(1 - c_p/X)}$. I again let

$$q_0 = \frac{\rho + \pi(1 - \rho) - c_p/X}{\rho(1 - c_p/X)}. \tag{24}$$
5.3.2 The Defendants’ Responses

Suppose that the prosecutor has offered $\mathbf{s} = \{s_1, s_2\}$ to the defendants. Each defendant’s strategy now depends not only on his beliefs about the prosecutor’s strategy but also on the other defendant’s strategy. If guilty, $D_i$ accepts a plea offer $s_i$ after observing $\mathbf{s}$ only if

$$s_i \leq (X + c_d)\hat{p}_{d_i}(t_i|r_i, r_j)\hat{p}_{d_i}(r_j) + (X + c_d)\hat{p}_{d_i}(t_i|r_i, a_j)\hat{p}_{d_i}(a_j).$$

(25)

The first term of the right hand side of inequality (25) is the expected sentence of $D_i$ given that both defendants reject, multiplied by the probability that the other defendant rejects. The second term is the expected sentence of $D_i$ given that he alone rejects, multiplied by the probability that the other defendant accepts.

Similarly, $D_i$ accepts a plea offer $s_i$ after observing $\mathbf{s}$ when innocent only if

$$s_i \leq (\pi X + c_d)\hat{p}_{d_i}(t_i|r_i, r_j)\hat{p}_{d_i}(r_j) + (\pi X + c_d)\hat{p}_{d_i}(t_i|r_i, a_j)\hat{p}_{d_i}(a_j).$$

(26)

**Lemma 5.1 (Symmetry).** If both defendants reject with positive probability, then acceptance probabilities are symmetric in equilibrium. Formally, if $\hat{p}_p(r_1), \hat{p}_p(r_2) > 0$, then $\hat{p}_p(r_1|g_1) = \hat{p}_p(r_2|g_2), \hat{p}_p(r_1|i_1) = \hat{p}_p(r_2|i_2)$

**Proof.** See Appendix A.

Lemma 5.1 reflects the intuition that if the defendants mix with different rates, then one defendant is more likely to be guilty post-rejection and becomes the first priority of the prosecutor. This cannot occur in equilibrium, as it then becomes the dominant strategy of the first priority defendant to accept all pleas $s < X + c_d$. Note that the symmetry holds even for asymmetric offers, so Lemma 5.1 implies that the prosecutor cannot induce one defendant to accept more frequently than the other unless one defendant accepts with probability 1.

**Lemma 5.2.** (Subgame equilibria) The prosecutor optimal subgame equilibria for a fixed
\( \bar{s} = \{s_1, s_2\} \) are as follows. For \( \bar{s} \) such that

(i) \( s_1, s_2 \leq \pi X + c_d \), both defendants accept their pleas. If a defendant deviates, the procurator tries.

(ii) \( s_1 \leq \pi X + c_d, s_2 \in (\pi X + c_d, X + c_d) \), \( D_1 \) accepts regardless of type, and \( D_2 \) rejects when innocent and mixes when guilty, accepting with probability \( q_0 \). The procurator tries \( D_1 \) if he rejects. If only \( D_2 \) rejects, the procurator tries \( D_2 \) with probability \( \frac{s_2}{X + c_d} \).

(iii) \( s_1 \leq \pi X + c_d, s_2 > X + c_d \), \( D_i \) accepts with probability 1 and \( D_2 \) rejects. The procurator tries \( D_1 \) if he deviates, and \( D_2 \) if else.

(iv) \( s_1, s_2 \in (\pi X + c_d, X + c_d] \) and \( s_1 + s_2 \leq (X + c_d)(1 + \rho q_0) \), defendants reject when innocent and accept with probability \( q_0 \) when guilty. The procurator mixes with some combination to make both defendants indifferent when guilty.\(^{23}\)

Both defendants reject their pleas with probability 1 for all other subgames.\(^{24}\)

**Proof.** See Appendix A. \( \square \)

In the procurator optimal subgame equilibria outlined in sub-lemmas (i), (ii), and (iii), \( D_1 \) pleads guilty with probability 1. This effectively isolates \( D_2 \), and \( D_2 \)'s strategies in these equilibria accordingly mirror those in the bilateral game. This result illustrates the negative bargaining externality between the defendants: although the procurator can only try one of the two defendants, the first defendant’s decision to plead guilty prevents the second from doing any better than he could if the procurator could try both defendants.

Notably, the procurator does best for a fixed \( \bar{s} \) when the negative externality between the two defendants is maximized in subgame equilibrium.

\(^{23}\)See Appendix A for details. For some \( \bar{s} \), the procurator can play an infinite number of strategies to satisfy these indifference requirements.

\(^{24}\)Except for the fact that in subgames for which \( s_1 \neq s_2, s_1 \) and \( s_2 \) can of course be switched without loss of generality.
Corollary 5.2.1. For a fixed $\mathbf{s} = \{s_1, s_2\}$, the prosecutor optimal subgame equilibrium is the subgame equilibrium that maximizes the expected number of acceptances.

Proof. See proof of Lemma 5.2 in Appendix.

Corollary 5.2.1 is consistent with the contracting with externalities literature, which has often shown the principal may profit by exploiting negative externalities between the defendants (Rasmusen et al., 1991; Segal and Whinston, 2000).

5.2.3 The Prosecutor’s Optimal Offer

Given that subgames are prosecutor optimal, one of three on-path strategies will occur in equilibrium.

Definition 5.1. I define three potential equilibria by their on-path strategies:

(i) Symmetric pooling equilibrium: the prosecutor offers $\mathbf{s} = \{\pi X + c_d, \pi X + c_d\}$, and both defendants accept with probability 1 regardless of type. The prosecutor tries if a defendant deviates.

(ii) Symmetric\textsuperscript{25} semi-separating equilibrium: the prosecutor offers $\mathbf{s}$ such that $s_1, s_2 < X + c_d$, $s_1 + s_2 = (1 + \rho q_0)(X + c_d)$. Both defendants reject with probability 1 when innocent and accept with probability $q_0$ when guilty. The prosecutor always tries a defendant when able. If both defendants reject, the prosecutor tries $D_1$ with probability $\frac{X + c_d - s_j}{(1 - \rho q_0)(X + c_d)}$.

(iii) Asymmetric equilibrium: the prosecutor offers $\mathbf{s} = \{\pi X + c_d, X + c_d\}$. $D_1$ accepts regardless of type. $D_2$ rejects when innocent and accepts when guilty with probability $q_0$. The prosecutor tries $D_1$ if he rejects, and tries $D_2$ if $D_1$ accepts and $D_2$ rejects.

\textsuperscript{25}To be clear, this equilibrium is symmetric because the strategies of the defendants are symmetric. The offers to the defendants need not be symmetric in this equilibrium.
The key intuitions of the three equilibria are as follows. In the symmetric pooling equilibrium, the prosecutor offers the maximum pleas that will be accepted with probability 1 regardless of the defendants’ types. In the symmetric semi-separating equilibrium, the prosecutor offers the maximum set of pleas that guilty types will accept with probability \( q_0 \). In the asymmetric equilibrium, the prosecutor offers a lower plea to one defendant to ensure that he accepts. Given this, she offers \( X + c_d \) to the other defendant, which is accepted by guilty types with probability \( q_0 \).

**Lemma 5.3 (Sequential Equilibrium).** Under the prosecutor optimal subgame refinement, if

(i) \( \pi X + c_d > \rho q_0(X + c_d) \), the symmetric pooling equilibrium is played.

(ii) if \( (\rho q_0)^2(X + c_d) > \pi X + c_d \), the symmetric semi-separating equilibrium is played.

(iii) if \( (\rho q_0)^2(X + c_d) < \pi X + c_d < \rho q_0(X + c_d) \), the asymmetric equilibrium is played.

**Proof.** See Appendix A.

Lemma 5.3 states that the prosecutor pursues one of three strategies in equilibrium. If the proportion of innocent defendants is high and innocent defendants have a relatively high expected disutility of trial, the prosecutor offers the maximum pleas that will be accepted by innocent defendants. If the proportion of innocent defendants is low and innocent defendants have a low expected disutility of trial, the prosecutor offers the maximum plea that guilty defendants will accept. Relative to the pooling equilibrium, the prosecutor extracts more utility from guilty defendants and less utility from innocent defendants in this semi-separating equilibrium. Finally, when parameters are middling, the prosecutor ensures the plea of one defendant and then offers a higher plea to the other defendant.
5.2.4 Welfare Analysis

The symmetric pooling equilibrium payouts are

\[ U_p = 2(\pi X + c_d), \quad U_g = -\pi X - c_d, \quad U_i = -\pi X - c_d. \] (27)

The symmetric semi-separating equilibrium payouts are

\[ U_p = \rho q_0(1 + \rho q_0)(X + c_d), \quad U_g = -\frac{X + c_d}{2}(1 + \rho q_0), \quad U_i = -\frac{\pi X + c_d}{2}(1 + \rho q_0). \] (28)

The asymmetric equilibrium payouts are

\[ U_p = (\pi X + c_d) + \rho q_0(X + c_d), \quad U_g = -\frac{(1 + \pi)X}{2} - c_d, \quad U_i = -\pi X - c_d. \] (29)

Note that in the symmetric pooling equilibrium and the asymmetric equilibrium, innocent defendants do as badly as possible, receiving their expected disutility of trial with probability 1. It is only in the symmetric semi-separating equilibrium that innocent defendants benefit from the prosecutor’s resource constraint. This result reflects the intuition that the defendants primarily benefit from the trial constraint when there is a non-zero probability that they will both reject their pleas.

<table>
<thead>
<tr>
<th></th>
<th>( U_g )</th>
<th>( U_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric pooling</td>
<td>(-\pi X - c_d)</td>
<td>(-\pi X - c_d)</td>
</tr>
<tr>
<td>Symmetric semi-separating</td>
<td>(-\frac{1}{2}(1 + \rho q_0)(X + c_d))</td>
<td>(-\frac{1}{2}(1 + \rho q_0)(\pi X + c_d))</td>
</tr>
<tr>
<td>Asymmetric</td>
<td>(-\frac{1}{2}(1 + \pi)X - c_d)</td>
<td>(-\pi X - c_d)</td>
</tr>
</tbody>
</table>

Table 5.1: Defendant welfare in different sequential equilibria
5.3 The Effect of Capped Plea Discounts

In both the symmetric pooling equilibrium and the asymmetric equilibrium, innocent defendants plead guilty with positive probability. Thus, capped plea discounts may prevent innocent defendants from pleading guilty in equilibrium. I assume that one can cap plea discounts to stop innocent defendants from pleading guilty without effectively banning plea bargaining.

**Assumption 5.2.** I assume that there exists some \( \bar{s} \) such that the defendants reject when innocent and accept with non-zero probability when guilty. For this to be true,

\[
\pi X + c_d < \frac{(X + c_d)(1 + \rho q_0)}{2}.
\]  

(30)

This assumption guarantees that capped plea discounts do not effectively ban plea bargaining. Although this inequality is mathematically somewhat opaque,\(^{26}\) the practical motivation behind the assumption is one of common sense. It would be very odd if innocent defendants and guilty defendants alike accepted below some threshold and rejected above the same threshold, implying that no plea offer could separate the types.

Given this assumption, I make the fundamental proposition of the paper.

**Proposition 5.3.** (i) If one restricts pleas such that \( s_1, s_2 > \pi X + c_d \), the unique sequential equilibrium of the game is the symmetric semi-separating equilibrium.

(ii) When binding, this restriction leaves innocent defendants strictly better off, has an ambiguous effect on the welfare of guilty defendants, and leaves prosecutors weakly worse off.

**Proof.** After the restrictions, there are no subgames available to the prosecutor in which innocent defendants will plead guilty. By Lemmas 5.2 and 5.3, the symmetric semi-
separating equilibrium becomes the unique sequential equilibrium of the game post-
restriction.

When binding, the restriction leaves the prosecutor weakly worse off, as she must
deviate from her first choice strategy. Because innocent defendants do strictly better in
the symmetric semi-separating equilibrium than in the other two sequential equilibria,
innocent defendants are strictly better off. If the symmetric pooling equilibrium was
played before the restriction, the reform leaves guilty defendants strictly worse off by
Assumption 5.2. If the asymmetric equilibrium was played before restriction, the effect
on guilty defendants is ambiguous (see Table 5.1 for welfare of defendants).

Proposition 5.3 provides the core result of the paper, showing that the introduction
of the trial constraint reverses the welfare effect that capped plea discounts have on
innocent defendants. The improvement for innocent types occurs because capped plea
discounts lead each defendant to reject with non-zero probability, decreasing the negative
externality between the defendants. When the prosecutor has the ability to induce pleas
from innocent types, she can fully exploit the negative externality between defendants
by ensuring that each defendant believes that he will be tried with probability 1 if he
rejects his plea. The maximum trial threat against each defendant is possible because the
prosecutor may offer low pleas to ensure that at least one defendant pleads guilty. Thus,
the defendants suffer from a coordination issue in these equilibria (symmetric pooling
and asymmetric), as they reject simultaneously with probability zero. Capped plea dis-
counts prevent the prosecutor from forcing this coordination issue between defendants by
ensuring that there are non-zero, independent probabilities that each defendant rejects
his plea in equilibrium. Then, the prosecutor can no longer make each defendant believe
that he will be tried with probability 1 if he rejects his plea, and her optimal strategy is to
offer higher pleas and focus on extracting large sentences from guilty types. Importantly,
the non-zero chance that both defendants reject their pleas particularly benefits inno-
cent types, as innocent types reject more frequently than guilty types in the post-reform
This improvement for innocent defendants would not necessarily be socially optimal if it were also accompanied by an improvement for guilty defendants. However, if the symmetric pooling equilibrium is played before the reform, capped plea discounts leave guilty defendants strictly worse off. This decrease in guilty defendant utility occurs because the effect of the increase in plea severity outweighs the effect of the decreased bargaining externality. Given the proportion of convictions secured through plea bargaining, there is reason to think that practices within the legal system resemble the symmetric pooling equilibrium, thus suggesting that capped plea discounts might be good social policy. Moreover, when the asymmetric equilibrium is played before the reform, the increase in plea severity is less drastic, but capped plea discounts may still leave guilty defendants worse off.

In contrast, while a full ban on plea bargaining would leave innocent defendants better off, it would also always leave guilty defendants better off. If pleas are banned, the prosecutor always tries one of the two defendants, yielding welfare outcomes

\[
U_p = \rho X + (1 - \rho)\pi X - c_p, \quad U_g = -\frac{X + c_d}{2}, \quad U_i = -\frac{\pi X + c_d}{2}.
\]

Although the change in \(U_s\) is ambiguous, the consensus that banning plea bargaining is not a viable option suggests that this result, in which half of all cases are dropped, is not socially desirable. The undesirability of the total ban is likely due to the large improvement for guilty types that it creates. Thus, the fact that capped plea discounts may help innocent defendants while hurting the guilty is a strong argument in favor of the reform.
5.4 Allowing Ex-Ante Commitment

I lastly analyze the game when the prosecutor may commit \textit{ex ante} to a trial strategy. When binding, capped plea discounts leave innocent defendants strictly better off and leave guilty defendants strictly worse off when the prosecutor can commit. Thus, the argument in favor of capped plea discounts is stronger when commitment is allowed, as there are no conditions in which capped plea discounts help guilty defendants.

**Proposition 5.4.** If the prosecutor can commit \textit{ex ante} to trial, she commits to trying $D_1$ and then $D_2$ if $D_1$ accepts, and then plays the one of the following three strategies:

1. The prosecutor offers $s = \{X + c_d, \rho(X + c_d)\}$. Guilty types accept and innocent types reject. This yields payout $U_p = \rho(1 + \rho)(X + c_d) + (1 - \rho^2)(\pi X - c_p)$.

2. The prosecutor offers $s = \{\pi X + c_d, X + c_d\}$. $D_1$ accepts regardless of type, and $D_2$ accepts if guilty and rejects if innocent. This yields payout $U_p = \pi X + c_d + \rho(X + c_d) + (1 - \rho)(\pi X - c_p)$.

3. The prosecutor offers $s = \{\pi X + c_d, \pi X + c_d\}$. Both defendants accept regardless of type. This yields payout $U_p = 2(\pi X + c_d)$.

**Proof.** See Appendix A.  

Committing \textit{ex ante} is a weakly dominant strategy for the prosecutor because it allows the prosecutor to prevent guilty types from mixing. Although the prosecutor commits to trying innocent defendants, she tries innocent defendants in equilibrium regardless of whether she commits.\footnote{Further intuition behind this is explained in the proof located in Appendix A.} The three possible equilibria with commitment mirror the three possible sequential equilibria of the game without commitment: the prosecutor either offers both defendants low pleas to ensure they both plead guilty, offers one defendant a low plea to isolate the second defendant, or offers both defendants high pleas.
Corollary 5.4.1. When the prosecutor can commit ex-ante to trial, restricting $s_1, s_2 > \pi X + c_d$ leaves guilty defendants weakly worse off and innocent defendants weakly better off.

Corollary 5.4.1 follows from the fact that the commitment strategy in which the prosecutor offers $\tilde{s} = \{X + c_d, \rho(X + c_d)\}$ becomes the prosecutor’s unique reasonable option. The welfare outcomes for the defendants from the three potential strategies can be seen in the table below.

<table>
<thead>
<tr>
<th>$\tilde{s}$</th>
<th>$U_g$</th>
<th>$U_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${X + c_d, \rho(X + c_d)}$</td>
<td>$\frac{1}{2}(1 + \rho)(X + c_d)$</td>
<td>$\frac{1}{2}(1 + \rho)(\pi X + c_d)$</td>
</tr>
<tr>
<td>${\pi X + c_d, \pi X + c_d}$</td>
<td>$\pi X + c_d$</td>
<td>$\pi X + c_d$</td>
</tr>
<tr>
<td>${\pi X + c_d, X + c_d}$</td>
<td>$\frac{1}{2}(1 + \pi)X + c_d$</td>
<td>$\pi X + c_d$</td>
</tr>
</tbody>
</table>

Table 5.2: Defendant welfare in equilibria when prosecutor commits

Innocent defendants are strictly better off when capped plea discounts bind because they are best off when the prosecutor offers the set of high pleas. Guilty defendants are strictly worse off when capped plea discounts bind because they are strictly worse off when the prosecutor offers the set of high pleas. This occurs because in the equilibrium forced by capped plea discounts, the prosecutor extracts the maximum sentence from guilty types at the cost of doing worse when defendants are innocent. Thus, when binding, capped plea discounts leave society strictly better, and there is evidence that capped plea discounts may help innocent defendants while hurting the guilty with or without commitment.

\(^{28}\)This inequality follows from Assumption 5.2.
Sequential Offers

While models of multilateral contracting frequently assume that offers occur simultaneously, many papers also consider games in which offers are made sequentially (Che and Spier, 2008; Che and Yi, 1993; Daughety and Reinganum, 2002; Hua and Spier, 2005). Although qualitative research suggests that prosecutors simultaneously handle multiple cases, it seems reasonable to assume that prosecutors leave some time in between offers (Alschuler, 1983). I show that the the main finding of the paper is stronger when offers are made sequentially. When the prosecutor is only able to try one defendant, restricting $s_1, s_2 > \pi X + c_d$ always binds and thus leaves innocent defendants strictly better off. The welfare effect of capped plea discounts on guilty defendants, however, is still ambiguous.

The timing of the sequential model is as follows. The prosecutor first offers a plea $s_1$ to $D_1$, after which $D_1$ may accept or reject the plea. All parties observe $D_1$’s response to the plea. After observing $D_1$’s response to the plea, the prosecutor offers a plea $s_2$ to $D_2$, who may accept or reject the plea. Finally, after both defendants have responded to their pleas, the prosecutor decides whether to try a defendant. The assumptions from earlier sections still hold.

It is clear that the timing of the offers does not affect the game when the prosecutor can try both defendants. This equivalency follows from the fact that the multilateral game is equivalent to two independent bilateral games when the prosecutor is not resource constrained. When two games are independent, the relative timing of the games does not matter. Thus, this section need only analyze the game when the prosecutor can try only one of the two defendants.

Lemma 6.1. When the prosecutor is able to try only one of two defendants and offers are made sequentially, there are two possible sequential equilibria:

(i) Symmetric pooling equilibrium: if $2(\pi X + c_d) \geq \rho q_0 (1 + \rho q_0) (X + c_d) + (1 - \rho q_0)(\pi X + c_d)$, the prosecutor offers $s_1 = \pi X + c_d$, which is accepted by $D_1$ with probability
1 regardless of type. The prosecutor then offers \( s_2 = \pi X + c_d \) regardless of the response of \( D_1 \), which is accepted by \( D_2 \) with probability 1 regardless of type.

(ii) Semi-separating dynamic equilibrium: if \( 2(\pi X + c_d) \leq \rho q_0 (1 + \rho q_0) (X + c_d) + (1 - \rho q_0)(\pi X + c_d) \): The prosecutor offers \( s_1 = X + c_d \) to \( D_1 \), who rejects when innocent and mixes when guilty, accepting with probability \( q_0 \). If \( D_1 \) rejects, the prosecutor offers \( s_2 = \pi X + c_d \) to \( D_2 \), who accepts with probability 1 regardless of type. If \( D_1 \) accepts, the prosecutor offers \( s_2 = \pi X + c_d \) to \( D_2 \), who rejects when innocent and mixes when guilty, accepting with probability \( q_0 \). The prosecutor tries \( D_2 \) if he rejects and \( D_1 \) if \( D_2 \) accepts and \( D_1 \) rejects.

The prosecutor is weakly better off and defendants are weakly worse off under these equilibria relative to equilibrium when offers are made simultaneously.

Proof. See Appendix A. \( \square \)

Lemma 6.1 illustrates that the prosecutor pursues one of two strategies in equilibrium. When the excepted sentence of innocent types is high and the proportion of innocent defendants is high, she offers the maximum pleas that will always be accepted by the defendants. This equilibrium mirrors the symmetric pooling equilibrium of the simultaneous game. When there is relatively more utility to be gained from targeting guilty types, she instead offers the maximum \( s_1 = X + c_d \) to \( D_1 \), and then adjusts her offer to \( D_2 \) based on \( D_1 \)'s response. Importantly, because she adjusts her offer to \( D_2 \) based on \( D_1 \)'s response, the prosecutor can ensure that at most one defendant pleads guilty. Removing the possibility that both defendants will plead guilty enables the prosecutor to threaten trial against both defendants with probability 1. While the prosecutor is also able to guarantee that at least one defendant pleads guilty in the asymmetric equilibrium of the simultaneous game, that equilibrium requires that she always offer a low plea to one defendant. When offers are sequential, on the other hand, she only needs to offer a low plea if \( D_1 \) rejects.
This additional flexibility leaves the prosecutor weakly better off in equilibrium relative to when offers are made simultaneously. The welfare outcomes of the symmetric pooling equilibrium are

$$U_p = 2(\pi X + c_d), \quad U_g = -\pi X - c_d, \quad U_i = -\pi X - c_d. \quad (32)$$

The welfare outcomes of the dynamic semi-separating equilibrium are

$$U_p = \rho q_0 (1 + \rho q_0)(X + c_d) + (1 - \rho q_0)(\pi X + c_d),$$
$$U_g = -\frac{1}{2}(1 + \rho q_0)(X + c_d) + \frac{1}{2}(1 - \rho q_0)(\pi X + c_d), \quad U_i = -\pi X - c_d. \quad (33)$$

In both possible equilibria, innocent defendants receive their maximum disutility, and guilty defendants are worse off relative to when offers are sequential. Note also that there is a non-zero chance that innocent defendants plead guilty in both equilibria, so the innocence problem always occurs in equilibrium when offers are sequential.

**Proposition 6.1.** In the sequential game, capping plea discounts such that $s_1, s_2 > \pi X + c_d$ is always binding. In the post-restriction equilibrium,

(i) the prosecutor offers $s_1 \in [\rho q_0(X + c_d), X + c_d] \cap (\pi X + c_d, X + c_d - (1 - \rho q_0)(\pi X + c_d))$.\footnote{The second set of the intersection is necessary to ensure that $s_1, s_2 > \pi X + c_d$. The upper bound of the second set is always lower than the upper bound of the first set. The first set’s upper bound is included to help illustrate the intuition that the $s_1 + s_2 \leq (1 + \rho q_0)(X + c_d)$.} $D_1$ rejects when innocent and accepts with probability $q_0$ when guilty. If $D_1$ rejects, the prosecutor offers $s_2 = \frac{X + c_d - s_1}{1 - \rho q_0}$ to $D_2$. If $D_1$ accepts, she offers $s_2 = X + c_d$. In both cases, $D_2$ rejects if innocent and accepts with probability $q_0$ if guilty. The prosecutor tries whenever a defendant rejects. If both defendants reject, the prosecutor tries $D_i$ with probability $\frac{X + c_d - s_j}{(1 - \rho q_0)X + c_d}$.

(ii) the prosecutor is strictly worse off, innocent defendants are strictly better off, and guilty defendants have an ambiguous change relative to the equilibrium before capped
plea discounts.

Proof. See Appendix A.

These capped plea discounts are always binding because both equilibria of the unregulated sequential game involve the prosecutor offering $s_i = \pi X + c_d$ with non-zero probability. Thus, after the restriction, equilibrium changes with probability 1.

The welfare outcomes of this game are identical to the welfare outcomes in the simultaneous game post-reform. This reflects the fact that the prosecutor’s advantage in the sequential game is neutralized when she cannot secure a plea with probability 1. The welfare outcomes are

$$U_p = \rho q_0 (1 + \rho q_0)(X + c_d), \quad U_g = -\frac{X + c_d}{2} (1 + \rho q_0), \quad U_i = -\frac{\pi X + c_d}{2} (1 + \rho q_0).$$

The prosecutor is unsurprisingly strictly worse off after this restriction, and innocent defendants are strictly better off because both equilibria before the restriction gave them their maximum disutility. The effect on guilty defendants is ambiguous. If the semi-separating dynamic equilibrium was played before the restriction, guilty defendants are strictly better off after the reform. On the other hand, if the symmetric pooling equilibrium was played before the restriction, guilty defendants are strictly worse off after the reform.

Thus, the finding that capped plea discounts may improve the welfare of innocent defendants without improving the welfare of guilty defendants is robust against sequential offers. If anything, the result is stronger, as capped plea discounts will always bind when the offers are made sequentially.
7 Conclusion

Given the centrality of plea bargaining to the legal system, a proper understanding of plea bargain reform is of high social importance. Building upon the work of Gazal-Ayal (2005) and Bar-Gill and Ben-Shahar (2009), I have demonstrated that limited prosecutorial resources have substantial theoretical implications for plea bargain reform. When prosecutors can try all of their cases, the traditional argument that capped plea discounts leave all parties worse off holds. However, when prosecutors face trial constraints, there is evidence that capped plea discounts may improve the welfare of innocent defendants while decreasing the welfare of guilty defendants.

While my model illustrates the theoretical implications of trial constraints, I suggest that it also has immediate practical implications for plea bargain reform. At a minimum, capped plea discounts should not be portrayed within the plea bargain debate as a reform that defies economic argument. Additionally, as reflected by the courts’ overwhelming reliance on plea bargaining, prosecutors are severely limited in their trial capacity. Thus, there is reason to think that introducing capped plea discounts might be effective social policy, for the theoretical conditions under which the reform may improve social welfare are met within the legal system.

That being said, further research is needed before any definitive claims can be made about the welfare implications of capped plea discounts. First, although this paper relaxes the major assumption of symmetric information in Bar-Gill and Ben-Shahar (2009), several strong assumptions remain. Future models could consider continuous types, generalize the model to \( N \) defendants, consider different risk preferences, or introduce secret offers. I offer a sketch of the analysis of the game under secret offers in Appendix B; however, I find that substantial assumptions about off-path beliefs are needed to make the game tractable. Second, further empirical and qualitative research must be done to better understand the strategies prosecutors use when plea bargaining. This research would provide evidence to guide both subgame equilibrium selection and modeling of
prosecutorial objective functions.
A Proofs

A.1 Section 4

Lemma 4.1 (Subgame equilibria in the bilateral game)

Proof. Fix \( s > X + c_d \). It is clear that the defendant prefers trial to settlement in this case.

Now fix \( s \in (0, X + c_d) \). The defendant cannot mix when innocent in equilibrium because if the he accepts with positive probability when innocent, then he must always accept when guilty. If the defendant never rejected when guilty, the prosecutor would never try, allowing all defendants to deviate by rejecting.

Thus, when innocent, the defendant must either play the pure strategy of accept or reject in equilibrium. Suppose that defendant rejects when innocent. Then, the guilty defendant must mix. If the guilty defendant always accepted, the prosecutor would never try. If the guilty defendants always rejected, the prosecutor would always try. Both of these states allow for the guilty defendant to deviate. Suppose the defendant accepts when innocent. Then, the guilty defendant must also accept.

Therefore, in subgame equilibrium for \( s \in (0, \pi X + c_d) \), the defendant must accept regardless of type or reject when innocent and mix when guilty. For the defendant to accept regardless of type, it must be that

\[
s \leq (\pi X + c_d)\tilde{p}_d^s(t|r), \tag{35}
\]

so this equilibrium type can exist for \( s \in (0, \pi X + c_d] \). If the defendant mixes when guilty, he must be indifferent when guilty. This indifference is characterized by the equation

\[
s = (X + c_d)p^s(t|r). \tag{36}
\]
For $\hat{p}_d^t(t|r) \in (0, 1)$ in equilibrium, the prosecutor must be indifferent between trial and dismissal, which can be characterized by

$$c_p = \frac{\pi \hat{p}^s(r|i)(1 - \rho) + \hat{p}^s(r|g)\rho X}{\hat{p}^s(r|i)(1 - \rho) + \hat{p}^s(r|g)\rho}.$$  \hspace{1cm} \text{(37)}

Given that innocent parties never accept, the unique solution to this equation is

$$\hat{p}^s(a|i) = 0$$  \hspace{1cm} \text{(38)}

$$\hat{p}^s(a|g) = \frac{\rho + \pi(1 - \rho) - c_p/X}{\rho(1 - c_p/X)}$$  \hspace{1cm} \text{(39)}

$$\hat{p}^s(t|r) = \frac{s}{X + c_d}$$  \hspace{1cm} \text{(40)}

Lemma 4.2 (Sequential equilibria in the bilateral game)

Proof. By Lemma 4.1, the prosecutor must select a plea that is accepted regardless of type, is rejected regardless of type, or leads to semi-separation. Among the pleas that are always accepted, it is clear that the maximum plea $s_A^*$ is optimal, yielding $U_p = s_A^*$. In the semi-separating subgames, the prosecutor is indifferent about going to trial, so her expected payoff comes solely from the acceptance the guilty pleas, and $U_p = s\rho q_0$. Thus, $s = X + c_d$ is optimal among these pleas, yielding payoff $U_p = \rho q_0(X + c_d)$. If a plea is always rejected, the prosecutor tries and receives utility $\rho X + (1 - \rho)\pi X - c_p$.

Note that $\rho q_0(X + c_d) > \rho X + (1 - \rho)\pi X - c_p$. This follows from Assumption 3.1. More intuitively, the prosecutor is better off when the guilty defendant accepts some of the time because she is able to avoid trial with non-zero probability, and the sentences of the defendants are the same regardless. Thus, it is always optimal for the prosecutor to offer $s = X + c_d$ or $s = s_A^*$, and Lemma 4.2 follows.\footnote{I assume that the semi-separating equilibrium is played in the subgame for $s = X + c_d$ because it}

\hspace{1cm} \square
A.2 Section 5

Proposition 5.1 (Subgame perfect equilibrium under perfect information)

Proof. Each defendant believes that the other defendant will accept with probability 1 when guilty. Thus, if a defendant deviates by rejecting when, he will be the only guilty defendant rejecting and will be tried. There is therefore no incentive for any party to deviate, as innocent defendants and the prosecutor do as well as possible and guilty defendants have no profitable deviations. □

Lemma 5.1 (Conditional symmetry of defendants’ strategies)

Proof. Suppose that \( \hat{p}^x_p(a_1|i_1) \neq \hat{p}^x_p(a_2|i_2) \). Without loss of generality, assume that \( \hat{p}^x_p(a_1|i_1) > \hat{p}^x_p(a_2|i_2) \). Then, \( D_1 \) accepts his plea with positive probability when innocent, meaning that he must always accept when guilty. It follows that if \( D_1 \) rejects his plea, the prosecutor never tries him because he is never guilty upon the rejection of a plea. This cannot occur in equilibrium, as \( D_1 \) could then profitably deviate by rejecting all non-zero pleas. Therefore, \( \hat{p}^x_p(a_1|i_1) = \hat{p}^x_p(a_2|i_2) \) in all subgame equilibrium if \( \hat{p}^x_p(r_1), \hat{p}^x_p(r_2) > 0 \).

Suppose that \( \hat{p}^x_p(a_1|g_1) \neq \hat{p}^x_p(a_2|g_2) \). Without loss of generality, assume that \( \hat{p}^x_p(a_1|g_1) > \hat{p}^x_p(a_2|g_2) \). Then, because \( D_1 \) and \( D_2 \) must accept with the same probability when innocent in subgame equilibrium, \( D_2 \) is more likely to be guilty that \( D_1 \) conditional upon rejection. Therefore, the prosecutor either tries \( D_2 \) whenever he rejects his plea or never tries either defendant. Neither of these strategies can occur in subgame equilibrium. If the prosecutor tries \( D_2 \) whenever he rejects is plea, then \( D_2 \) has the weakly dominant strategy of accepting any \( s_2 \leq X + c_d \) in equilibrium when he is guilty. This contradicts the supposition that \( \hat{p}^x_p(a_1|g_1) \neq \hat{p}^x_p(a_2|g_2) \). If \( s_2 > X + c_d \), then the \( D_2 \) rejects and is always tried. \( D_1 \) can then deviate by always rejecting, a contradiction. If the prosecutor is the unique equilibrium for \( s = X + c_d - \epsilon \) for arbitrarily small \( \epsilon \). This minor assumption does not change equilibrium analysis, as if the defendant accepted less frequently at \( s = X + c_d \), the prosecutor could simply offer \( s = X + c_d - \epsilon \).
never tries either defendant, then both defendants should reject any non-zero plea, also a contradiction. Therefore, \( \hat{p}_p(a_1|g_1) = \hat{p}_p(a_2|g_2) \) if \( \hat{p}_p(r_1), \hat{p}_p(r_2) > 0 \). 

Lemma 5.2 (Prosecutor optimal subgame equilibria)

Proof. I first show that the complete set of subgame equilibria as follows. For \( \tilde{s} \) s.t.

1. \( s_1, s_2 \leq \pi X + c_d \), there exists a subgame equilibrium in which both defendants accept their plea regardless of their type.

2. \( s_1 + s_2 \geq X + c_d \), there exists a subgame equilibrium in which both defendants reject their pleas regardless of their type. The prosecutor tries one of the defendants upon the rejections, but is indifferent as to whom.

3. \( s_1 + s_2 \in (X + c_d, (X + c_d)(1 + \rho q_0)) \), \( s_1, s_2 \leq X + c_d \), there exists a subgame equilibrium in which innocent defendants reject their pleas and guilty defendants accept with probability \( \frac{(s_1 + s_2)/(X + c_d) - 1}{\rho} \).

4. \( s_1 + s_2 \in [0, (X + c_d)(1 + \rho q_0)] \), \( s_1, s_2 \leq X + c_d \), there exists a subgame equilibrium in which which innocent defendants reject their pleas and guilty defendants accept with probability \( q_0 \).

5. \( s_i \in [0, \pi X + c_d], s_j \in [\pi X + c_d, X + c_d] \), there exists a subgame equilibrium in which \( D_i \) accepts with probability 1 and \( D_j \) accepts with probability \( q_0 \).

6. \( s_i \in [0, \pi X + c_d], s_j \geq X + c_d \), there exists a subgame equilibrium in which \( D_i \) accepts and \( D_j \) rejects.

To prove that these equilibria are exhaustive, I first consider the 9 types of possible symmetric equilibria.

(i) Guilty mix, innocent accept: cannot occur because if the guilty are indifferent, the innocent prefer rejection.
(ii) Guilty mix, innocent mix: cannot occur because if the guilty are indifferent, the innocent prefer rejection.

(iii) Guilty mix, innocent reject: for this to occur in equilibrium the following condition must hold:

\[
\frac{s_i}{X + c_d} = \hat{p}_{a_i}^s(t_i|a_j) \left[ \hat{p}_{a_i}^s(a_j|g_j) \rho + \hat{p}_{a_i}^s(a_j|i_j)(1 - \rho) \right] + \\
\hat{p}_{a_i}^s(t_i|r_j, r_i) \left[ \hat{p}_{a_i}^s(r_j|g_j) \rho + \hat{p}_{a_i}^s(r_j|i_j)(1 - \rho) \right]
\]

(41)

Because the innocent reject in these equilibrium, this condition can be simplified to

\[
\frac{s_i}{X + c_d} = \hat{p}_{a_i}^s(t_i|r_i, a_j) \left[ \hat{p}_{a_i}^s(a_j|g_j) \rho \right] + \hat{p}_{a_i}^s(t_i|r_i, r_j) \left[ \hat{p}_{a_i}^s(r_j|g_j) \rho + (1 - \rho) \right]
\]

(42)

There are two solutions to these equations. In the first solution, the prosecutor is not indifferent about going to trial and tries whenever possible:

\[
\hat{p}_{a_i}^s(a_1|g_1) = \hat{p}_{a_i}^s(a_2|g_2) = \frac{(s_1 + s_2)/(X + c_d) - 1}{\rho}
\]

(43)

\[
\hat{p}_{a_i}^s(t_i|r_i, a_j) = 1
\]

(44)

\[
\hat{p}_{a_i}^s(t_i|r_i, r_j) = \frac{X + c_d - s_j}{2(X + c_d) - s_i - s_j}
\]

(45)

This equilibrium can exist for \(s_1 + s_2 \in [X + c_d, (X + c_d)(1 + \rho q_0)]\). In the second, the prosecutor is indifferent about proceeding to trial. One possible set is:

\[
\hat{p}_{a_i}^s(a_1|g_1) = \hat{p}_{a_i}^s(a_2|g_2) = q_0
\]

(46)


\[
\begin{align*}
\hat{p}_{d_1}(t_1|r_1, a_j) &= \frac{2s_i}{(X + c_d)(1 + \rho q_0)} \\
\hat{p}_{d_2}(t_1|r_1, r_j) &= \frac{2s_i}{(X + c_d)(1 + \rho q_0)} \frac{X + c_d - s_j}{(1 - q_0)(X + c_d)}
\end{align*}
\]

(iv) Guilty accept, innocent reject: cannot occur. Suppose not. Then, the prosecutor would never try, allowing guilty types to deviate by rejecting.

(v) Guilty accept, innocent mix: same as (iv).

(vi) Guilty accept, innocent accept: If a defendant accepts when innocent, he must accept when guilty. Thus, this equilibrium exists for all \( \bar{s} \) s.t.:

\[
s_i \leq \pi X + c_d
\]

(vii) Guilty reject, innocent accept: cannot exist. If innocent types accept with positive probability, guilty types must also accept.

(viii) Guilty reject, innocent mix: cannot exist (see (vii)).

(ix) Guilty reject, innocent reject. This equilibrium holds if

\[
s_i \geq (X + c_d)\hat{p}_{d_2}(t_1|r_1) \tag{50}
\]

If both parties reject, then \( \hat{p}_{d_1}(t_1|r_1) + \hat{p}_{d_2}(t_2|r_2) = 1 \), so this can exist for \( \bar{s} \) s.t. \( s_1 + s_2 \geq 1 \).

I now consider asymmetric equilibria. By Lemma 5.1, asymmetric subgame equilibria can occur only when one defendant accepts with probability 1. This may occur in equilibrium if the prosecutor has a high enough chance of being guilty. If only the guilty type trembles, then a belief that the defendant is always guilty after rejecting is possible in sequential equilibrium. Without loss of generality, assume that \( D_1 \) accepts with prob-
ability 1. Then, $D_2$ is effectively isolated, and the equilibrium mirrors that in the single defendant game. It follows that the aforementioned list is exhaustive.

I now show that that the prosecutor optimal outcomes are those described in Lemma 5.2. I refer to the possible subgames by the number in the “exhaustive list.”

(i) Fix $\bar{s}$ s.t. $s_1, s_2 \leq \pi X + c_d$. Subgame equilibrium is then either (1) or (4). In (4), her expected utility is $(s_1 + s_2)\rho q_0$ because she is indifferent about trial. In (1), her utility is $s_1 + s_2$. So, (1) is optimal for this range.

(ii) Fix $\bar{s}$ s.t. $s_1, s_2 > \pi X + c_d$, s.t. $s_1 + s_2 \leq (X + c_d)(1 + \rho q_0)$. Subgame equilibrium is then (2), (3), (4). (2) is dominated by (4). Note that (2), the pure rejection, can only exist when $s_1 + s_2 > X + c_d$. The prosecutor’s utility for (4) again is $(s_1 + s_2)\rho q_0$. At a minimum over the range (2), this is equal to $\rho q_0(X + c_d) > \rho(X + c_d) + (1 - \rho)\pi X - c_p$. This inequality can be shown algebraically using Assumption 3.1 and also follows from the equilibrium analysis of the bilateral game.

I now show that (4) also dominates (3). The prosecutor’s utility from (3) is

$$U_p = \frac{(s_1 + s_2)/(X + c_d) - 1}{\rho}(s_1 + s_2) + \frac{(s_1 + s_2)/(X + c_d) - 1}{\rho}(1 - \frac{(s_1 + s_2)/(X + c_d) - 1}{\rho})(s_1 + s_2 + T(.)) + (1 - \frac{(s_1 + s_2)/(X + c_d) - 1}{\rho})^2 T(.) \tag{53}$$

where $T(.)$ is the value of trial as a function of $s_1$ and $s_2$. It can be shown that $U_p$ is a function on $s_1$ and $s_2$ only in relation to how it is a function of $s_1 + s_2$. Moreover, $\frac{\partial U_p}{\partial (s_1 + s_2)} > 0$ and $\frac{\partial^2 U_p}{\partial (s_1 + s_2)^2} > 0$.

Note that at the lower end of the range of (3), $s_1 + s_2 = X + c_d$, (3) yields payout zero, and (4) has positive payout. Moreover, note that (3) and (4) yield the same payout at $s_1 + s_2 = (1 + \rho q_0)(X + c_d)$. Finally, note that (4)’s payouts are linear in
It follows that the payouts from (4) are greater than the payouts from (3) for the relevant range of pleas.

(iii) Fix \( s \) s.t. \( s_1 \leq \pi X + c_d, s_2 \in (\pi X + c_d, X + c_d] \). The possibilities are (2), (3), (4), (5). As shown above, (4) dominates (3) and (2). (5) dominates (4), as it has payout \( s_1 + \rho q_0 s_2 > \rho q_0 (s_1 + s_2) \). By transitivity, (5) is optimal for the prosecutor.

(iv) Fix \( s \) s.t. \( s_1 \leq \pi X + c_d, s_2 > X + c_d, D_1 \). Then, (2) and (6) are possible. Clearly, since \( D_2 \) always rejects, it is better to get \( D_1 \) to plead guilty, so (6) is best for the prosecutor.

Lemma 5.2 is finally proven. Corollary 5.2.1 is also proven, as the prosecutor optimal equilibria are all the options in which the defendants accept with highest probability.

**Lemma 5.3 (Sequential equilibria when prosecutor is resource constrained)**

*Proof.* Lemma 5.3 quickly follows from Lemma 5.2. Consider the optimal plea offers within each of the sub-lemmas, (i), (ii), (iii), and (iv)

(i) If pleas are accepted with probability 1, the prosecutor clearly chooses the largest pleas, and thus chooses \( s = \{ \pi X + c_d, \pi X + c_d \} \), yielding \( U_p = 2(\pi X + c_d) \).

(ii) Since \( D_1 \) accepts, the prosecutor offers \( s_1 = \pi X + c_d \). Because, \( D_2 \) mixes to make the prosecutor indifferent, she only gets expected utility from inducing pleas, and thus offers \( s_2 = X + c_d \) to \( D_2 \). This yields \( U_p = \pi X + c_d + \rho q_0 (X + c_d) \).

(iii) Since \( D_1 \) accepts, the prosecutor offers \( s_1 = \pi X + c_d \). Because \( D_2 \) rejects, the prosecutor can offer any \( s_2 > X + c_d \) to \( D_2 \). This yields \( U_p = \pi X + c_d + \rho X + (1 - \rho) \pi X - c_p \).

(iv) The prosecutor is indifferent about trial and thus gets expected utility only from pleas. She thus offers \( s \) to maximize the sum of \( s_1 \) and \( s_2 \), and receives utility \( \rho q_0 (1 + \rho q_0) (X + c_d) \).
Note that the best outcome in (iii) is strictly less than the best outcome in (iv). This is clear from the fact that, when all parties know that $D_1$ will plea out, the bargaining with $D_2$ mirrors the bilateral game. Lemma 5.3 follows, as the prosecutor selects the optimal plea from the sub-lemmas (i), (ii), and (iv).

Proposition 5.4 (Commitment outcome when prosecutor is resource constrained)

Proof. Suppose the prosecutor commits to trying $D_1$ and then $D_2$ if $D_1$ rejects. Then, $D_1$ will accept any $s_1 \leq X + c_d$ if guilty and will accept $s_1 \leq \pi X + c_d$. Thus, it is clear that the prosecutor will offer either $s_1 = X + c_d$ or $s_1 = \pi X + c_d$.

If the prosecutor offers $s_1 = X + c_d$, then $D_2$ knows that he will be tried if and only if $D_1$ is guilty, so he accepts $s_2 \leq \rho(X + c_d)$ if guilty and $s_2 \leq \rho(\pi X + c_d)$ if innocent. Thus, the prosecutor clearly offers either $s_2 = \rho(X + c_d)$ or $s_2 = \rho(\pi X + c_d)$.

If the prosecutor offers $s_1 = \pi X + c_d$, then $D_2$ knows that he will be tried, so he accepts if the plea is less than or equal to his expected sentence. Again, the prosecutor should therefore offer either $s_2 = X + c_d$ or $s_2 = \pi X + c_d$.

Thus, the reasonable strategies are $\bar{s} = \{X + c_d, \rho(X + c_d)\}$, $\{X + c_d, \rho(\pi X + c_d)\}$, $\{\pi X + c_d, \pi X + c_d\}$, or $\{\pi X + c_d, X + c_d\}$. However, $\{\pi X + c_d, X + c_d\}$ is preferable to $\{X + c_d, \rho(\pi X + c_d)\}$. The intuition behind this is that trial is guaranteed for both defendants in the first strategy and not the second. The three other strategies may be optimal, depending on the parameters of the model, as stated in the proposition.

Note that committing weakly dominates not committing. This reflects the fact sequential equilibria either involve defendants pleading guilty with probability 1 or mixing. The prosecutor can still secure pleas with probability 1 under commitment, but the edge from committing comes when guilty parties would mix in equilibrium without commitment. Then, her commitment allows her to prevent the guilty types from mixing, giving

\[\text{Note that committing to try a defendant with probability } 1 - \gamma \text{ simply scales the rent from that defendant } 1 - \gamma.\]
her the full sentence without having to go to trial. Although she must commit to try-
ing the innocent, which has negative expected value, she will paying the cost of trial
regardless of whether she commits, as innocent types always reject. □

A.3 Section 6

Lemma 6.1 (Sequential equilibrium under sequential offers)

Proof. I solve the game using backwards induction. I fix \( s_1, s_2 \in (0, X + c_d] \). In the
last stage of the game, the prosecutor decides whether to proceed to trial. As in the
simultaneous game, she tries the defendant who is more likely to be guilty provided that
the defendant is likely enough to be guilty that trial has positive expected value.

Next, consider any subgame in which \( D_1 \) has accepted his plea. Then, the subgame is
analogous to the bilateral game. By Lemma 4.2 and the prosecutor optimal refinement,
the subgame equilibrium is one the the following two equilibria.

(i) Pooling equilibrium: If \( \pi X + c_d \geq \rho q_0(X + c_d) \), the prosecutor offers
\( s_2 = \pi X + c_d \),
which the defendant accepts regardless of type.

(ii) Semi-separating equilibrium: If \( \pi X + c_d \leq \rho q_0(X + c_d) \), the prosecutor offers \( s_2 =
X + c_d \), which the defendant accepts regardless of type.

Now, consider the subgame in which \( D_1 \) has rejected his plea \( s_1 \) and the prosecutor has
offered \( s_2 \) to \( D_2 \). \( D_2 \)’s decision is very similar to his decision in the simultaneous offer
game. The main difference is that he knows with probability 1 that \( D_1 \) has rejected. \( D_2 \)
accepts when guilty only if

\[
s_2 \leq (X + c_d)\hat{p}_{d_2}(t_2|r_1, r_2)
\]  

(54)

and accepts when innocent only

\[
s_2 \leq (\pi X + c_d)\hat{p}_{d_2}(t_2|r_1, r_2).
\]  

(55)
Now consider decision of $D_1$. $D_1$’s decision is parallel to his decision in the simultaneous game. Let $\hat{p}_{d_1}^{s_1}$ denote $D_1$’s belief given fixed $s_1$. He accepts when guilty only if

$$s_1 \leq (X + c_d)\hat{p}_{d_1}^{s_1}(t_1|r_1).$$

(56)

The lemma of symmetry of mix rates given non-zero chance of rejection still holds. Note that for $D_1$ to mix, he must be tried with probability between 0 and 1. Thus, when $D_1$ mixes, it must be that $D_2$ is mixing with the same rate. The indifference equations are:

$$s_1 = (\pi X + c_d)[\hat{p}_{d_1}^{s_1}(t_1|r_1, r_2)\hat{p}_{d_2}^{s_1}(r_2|r_1) + \hat{p}_{d_1}^{s_1}(t_1|r_1, a_2)\hat{p}_{d_2}^{s_1}(a_2|r_1)].$$

(57)

$$s_2 = (X + c_d)\hat{p}_{d_2}^{s_2}(t_2|r_1, r_2).$$

(58)

Solving these indifference equations reveals that the possible defendant mix rates are the same as those in the simultaneous game. As shown in the proofs of Section 5, the best mixed equilibrium for the prosecutor is that in which the innocent defendant reject and guilty defendants accept with probability $q_0$. This outcome is also better than the outcome in which both defendants reject, also proven in the proofs for Section 5.

Thus, the best symmetric subgame path (with positive rejection rate) for the prosecutor for fixed $s_1 \in (\pi X + c_d, X + c_d)$, $s_1 + s_2 \in (X + c_d, (1 + \rho q_0))$ is for the $D_1$ to reject when innocent and accepts with probability $q_0$ when guilty. If $D_1$ rejects, fix $s_2$. Then, $D_2$ rejects when innocent and accepts with probability $q_0$ when guilty. The prosecutor tries at a rate that makes both defendants indifferent, as in the simultaneous game. After some messy algebra, the prosecutor’s payoff from this strategy is

$$U_p = \rho q_0(s_1 + s_2)$$

(59)

so her maximum payout from a symmetric subgame in which $D_1$ rejects is $U_p = \rho q_0(1 + \rho q_0)(X + c_d)$. 

57
The asymmetric equilibria are slightly different. Again, by symmetrical argument of Lemma 5.1, asymmetric equilibria require one defendant to plead guilty with probability 1. If $D_1$ pleads guilty with probability 1, the prosecutor’s best path is clearly $s_1 = \pi X + c_d$ and then the equilibrium of the bilateral game. If $D_2$ pleads guilty with probability 1, then the best path of the prosecutor is for $D_1$’s bargain to be the bilateral game and for $s_2 = \pi X + c_d$. When the bilateral game equilibrium is the pooling equilibrium, it is clear that it is best for the prosecutor to offer both defendants $\pi X + c_d$ and for both defendants to accept. Thus, the interesting case is when semi-separating equilibrium is the payout of the bilateral game. Then, payouts of these strategies are

$$U_p = \pi X + c_d + \rho q_0(X + c_d)$$  \hspace{1cm} (60)$$

Note, however, that the prosecutor can ensure that one defendant pleads guilty by offering a low plea to $D_2$ only if $D_1$ rejects his high plea. This strategy, which is described in detail in Lemma 6.1, yields payout

$$U_p = \rho q_0(1 + \rho q_0)(X + c_d) + (1 - \rho q_0)(\pi X + c_d)$$ \hspace{1cm} (61)$$

which dominates the payouts from the symmetric subgames as well as the other asymmetric subgames. Thus, the best strategy for the prosecutor (assuming prosecutor optimal subgames) is either

1. The prosecutor offers $\pi X + c_d$ to both defendants, and both defendants always accept. This yields payout $2(\pi X + c_d)$ for the prosecutor.

2. The prosecutor offers $s_1 = X + c_d$. $D_1$ rejects when innocent and mixes when guilty, accepting with rate $q_0$. If $D_1$ accepts, the prosecutor offers $X + c_d$ to $D_2$, to which $D_2$ responds identically to how $D_1$ responds. If $D_1$ rejects, the prosecutor offers $s_2 = \pi X + c_d$, which is accepted. This yields $U_p = \rho q_0(1 + \rho q_0)(X + c_d) + (1 - \rho q_0)(\pi X + c_d)$.  

58
and Lemma 6.1 follows.

**Proof of Proposition 6.1 (Effect of capped plea discounts when offers are sequential)**

*Proof.* When \( s_i > \pi X + c_d \), the unique equilibrium of the subgame when \( D_1 \) rejects
is clearly the unique subgame post-restriction of the bilateral game. What happens in
other cases is more tricky. Innocent defendants must reject all pleas in equilibrium post
restriction. Thus, guilty types must either reject or mix. If \( D_1 \) rejects, then \( D_2 \) also
rejects in equilibrium by the aforementioned symmetry, which is clearly a sub-optimal
outcome for the prosecutor. As shown in an earlier proof, when defendants mix, it is
best for the prosecutor if they reject when innocent and accept with probability \( q_0 \) when
guilty. For this mixing to occur, the following indifference equations must hold:

\[
\begin{align*}
    s_1 &= (\pi X + c_d)[\hat{p}_{d_1}^{s_1}(t_1|r_1, r_2)\hat{p}_{d_1}^{s_1}(r_2|r_1) + \hat{p}_{d_1}^{s_1}(t_1|r_1, a_2)\hat{p}_{d_1}^{s_1}(a_2|r_1)]. \\
    s_2 &= (X + c_d)\hat{p}_{d_2}^{s_2}(t_2|r_1, r_2).
\end{align*}
\]

These are the indifference equations of the two defendants. The prosecutor is also indiffer-
ent about trial because the defendants reject when innocent and accept with probability
\( q_0 \) when guilty. Because the prosecutor is indifferent about trial, her expected rent comes
from guilty pleas. Because the rate of mixing does not change with the specific pleas
offer, she offers the pleas that maximize her expected rent from pleas and can also sustain
these indifference equations. Clearly, the optimal equilibrium will be one in which she
always tries despite her indifference to exert maximal pressure of the defendants. When
she always tries, the solution to the defendants’ indifference equation is

\[
s_2 = \frac{X + c_d - s_1}{1 - \rho q_0}
\]
Given that she always tries, her utility as a function of \(s_1\) and \(s_2\) can be written

\[ U_p = \rho q_0 (s_1 + X + c_d) + (1 - \rho q_0) \left( \rho q_0 \frac{X + c_d - s_1}{1 - \rho q_0} \right) \]  

which simplifies to

\[ U_p = (1 + \rho q_0)(X + c_d) \]  

so the prosecutor is indifferent across these options. The proposition follows.

\[ \square \]
B Secret Offers

The contracting literature illustrates that when offers are secret, equilibrium may shift substantially relative to when offers offers are observable (Miklos-Thal and Shafer, 2016). I now suppose that each defendant observes his own plea offer but not the plea offer made to the other defendant. This appendix is a sketch for future research, as my analysis requires extensive assumptions and does not provide substantial intuition. Mainly, I show that analysis of capped plea discounts is degenerate under my refinement of passive beliefs, as capping plea discounts prevents equilibrium from occurring given my assumptions. I suggest that different refinements are therefore necessary to properly analyze the problem.

The literature on unobservable contracts illustrates that reasonable restrictions must be put on out-of-equilibrium beliefs for equilibrium to be tractable (Hart et al., 1990; Miklos-Thal and Shafer, 2016). This result follows from the fact that when an agent receives an off-equilibrium offer, rationality places no restrictions on his beliefs about the offers the other parties received. The most common refinement is the assumption that agents have “passive beliefs.” When agents have passive beliefs and receive an unexpected offer (one that deviates from the “candidate equilibrium”), their beliefs about the offers to other agents do not change. In other words, agents receiving an off-equilibrium offer believe that all other agents receive their equilibrium offers and act accordingly (Hart et al., 1990; McAfee and Schwartz, 1994).

In multilateral contracting games in which the principal does not act after making her offers, the refinement of passive beliefs is often sufficient to determine the agents’ decisions. In this paper, however, because the game does not end after agents’ actions, further refinements about the continuation game after the deviation are needed to characterize equilibrium in a sensible manner. I assume that all parties believe that when the prosecutor deviates from the candidate equilibrium in his offer to a defendant, the continuation game is in equilibrium. This paper is one of the first to analyze a game
in which an informed principal acts after the agents’ responses under secret offers, so this additional refinement is not established in the literature. I suggest that different refinements may be an avenue for further research on this type of multilateral contracting under secret offers.

**Proposition B.1.** When defendants have passive beliefs, off-path continuation games are in bilateral equilibrium, and offers are secret, defendants must accept with probability 1 in equilibrium. Thus, no equilibrium exists after restricting $s > \pi X + c_d$ under these refinements, and the analysis is degenerate.

**Proof.** Suppose that $\rho q_0(X + c_d) > \pi X + c_d$. In the candidate equilibria, at least one defendant is offered a plea less than $X + c_d$. Without loss of generality, assume that this defendant is $D_1$. The prosecutor can always profitably deviate by offering $s_1 = X + c_d$. This follows from the assumption that the continuation game between $D_1$ and the prosecutor is in equilibrium. The prosecutor then gets expected sentence $\rho q_0(X + c_d)$ from the defendant, which is the best outcome for the prosecutor if $\rho q_0(X + c_d) > \pi X + c_d$, as it is the utility she can achieve in the bilateral game. Thus, there is no equilibrium under these refinements when $\rho q_0(X + c_d) > \pi X + c_d$.

Suppose instead that $\rho q_0(X + c_d) \leq \pi X + c_d$. Then, if in equilibrium, the defendants are both offered and accept $\pi X + c_d$, equilibrium can hold because the prosecutor is already doing optimally. However, after the restriction, this equilibrium cannot occur. When the prosecutor cannot secure pleas, it is again optimal for her to offer to deviate from any candidate equilibrium $X + c_d$ and extract maximum sentence from each defendant without going to trial. The proposition follows.

Rather than illustrating that the model is not robust to passive beliefs, I suggest that this proposition illustrates the difficulty of refining a game in which the principal acts again after deviating. While I believe that the assumption that the continuation game is in equilibrium is an intuitive one, other refinements may be preferable to avoid degeneracy and are an avenue for future reference.
C Game Theory Definitions

This appendix defines the game theory concepts necessary to understand this paper. These definitions are adapted from the lecture notes for Economics 14.12, a game-theory course taught at the Massachusetts Institute of Technology by Professor Muhamet Yildiz. I am indebted to Professor Yildiz for letting me use these definitions.

**Definition C.1** (Nash equilibrium). A strategy profile \( s^* = (s_1^*, ..., s_n^*) \) is a Nash Equilibrium if and only if

\[
u_i(s_1^*, ..., s_{i-1}^*, s_i^*, s_{i+1}^*, ..., s_n^*) \geq u_i(s_1^*, ..., s_{i-1}^*, s_i, s_{i+1}^*, ..., s_n^*)
\]

for all \( s_i \) for all \( i \).

**Definition C.2** (Subgame perfect Nash equilibrium). A Nash equilibrium is subgame perfect if and only if it is a Nash equilibrium in every subgame of the game.

**Definition C.3** (Belief assessment). A belief assessment is a list \( \hat{p} \) of probability distributions on information sets; for each information set \( I \), \( \hat{p} \) gives a probability distribution \( \hat{p}(\cdot|I) \) on \( I \).

**Definition C.4** (Sequential rationality). For a given pair \((s, \hat{p})\) of strategy profile \( s \) and belief assessment \( \hat{p} \), strategy profile \( s \) is said to be sequentially rationally if and only if, at each information set \( I \), the player who is to move at \( I \) maximizes his utility

1. given his beliefs \( \hat{p}^*(\cdot|I) \) at the information set (which imply that he is at information set \( I \)), and

2. given that the players will play according to \( s \) in the continuation game.

**Definition C.5** (Consistency). Given any \((s, \hat{p})\), belief assessment \( \hat{p} \) is consistent with \( s \) if and only if there exist some trembling probabilities that go to zero such that the conditional probabilities derived by Bayes rule with trembles converge to probabilities given by \( \hat{p} \) on
all information sets (on and off the path of $s$). That is, there exists a sequence $(s^m, \hat{p}^m)$ of assessments such that

1. $(s^m, \hat{p}^m) \rightarrow (s, \hat{p})$, 

2. $s^m$ is “completely mixed” for every $m$, and

3. $\hat{p}^m$ is derived from $s^m$ using Bayes’ rule

**Definition C.6** (Sequential equilibrium). A pair $(s, \hat{p})$ of a strategy profile $s$ and a belief assessment $\hat{p}$ is said to be a sequential equilibrium if and only if $(s, \hat{p})$ is sequentially rational and consistent with $s$. 
References


McCoy, Candace (2005), “Plea bargaining as coercion: The trial penalty and plea bargaining reform.” Crim. LQ.


