American Law and Economics Review Advance Access published October 12, 2007

Guilt Shall Not Escape or Innocence Suffer? The Limits of Plea Bargaining When Defendant Guilt is Uncertain

David Bjerk, *Department of Economics, Claremont McKenna College*

This article examines optimal prosecutor behavior with respect to plea bargaining when defendant guilt is uncertain. I show that when jury beliefs and behavior are determined endogenously in equilibrium along with defendant and prosecutor behavior, plea bargaining can play only a limited role in managing society's conflicting desires to maximize punishment of the guilty and minimize punishment of the falsely accused. In particular, while it can be optimal for prosecutors to use plea bargaining to induce a large fraction of guilty defendants to voluntarily sort themselves from the innocent, such sorting must come at the cost of imposing relatively short sentences on such guilty defendants who accept plea bargains.

"The United States Attorney is the representative not of an ordinary party to a controversy but of a sovereignty whose obligation to govern impartially is as compelling as its obligation to govern at all; and whose interest, therefore, in a criminal prosecution is not that it shall win a case, but that justice shall be done.

Thanks to Phil Curry, Abigail Payne, Jennifer Reinganum, and participants at the 2006 ALEA conference for very helpful comments on earlier versions of this paper. JEL codes K14, K41, D81. A previous version of this paper was entitled "Guilt Shall Not Excape or Innocence Suffer: A Theory of Optimal Prosecutor Behavior When Defendant Guilt is Uncertain", but was substantially different from this current one.

Send correspondence to: David Bjerk, Department of Economics, Claremont McKenna College, 500 E. Ninth St., Claremont, CA 91711, USA; E-mail: david.bjerk@gmail.com.

American Law and Economics Review
doi:10.1093/aler/ahm010

As such, he is in a peculiar and very definite sense the servant of the law, the twofold aim of which is that guilt shall not escape or innocence suffer." *Supreme Court opinion in Berger v. United States* (as quoted by Reinganum (1988)).

1. Introduction

As long as it is possible that innocent individuals can be accused of crimes they did not commit, it may not always be possible to determine a given defendant's guilt with certainty. This means that society will always face a tension between wanting to decrease the standard of evidence necessary for conviction and increase sentence lengths in order to increase the expected punishment of the guilty, while simultaneously wanting to increase the standard of evidence necessary for conviction and decrease sentence lengths in order to minimize the expected punishment of those falsely accused.

It has been suggested that plea bargaining may provide a potential way of managing this tension (Grossman and Katz (1983), Reinganum (1988), and Baker and Mezzetti (2001)). To understand how, first note that length of a plea bargain sentence a given defendant is willing to accept will depend on his expected probability of being convicted at trial, which in turn will generally depend on two variables: (i) the strength of the initial case against him, and (ii) whether or not he is actually guilty. The first variable may affect a defendant's expected probability of conviction at trial if actual guilt generally leads to a strong initial case against a defendant, and if a strong initial case generally leads to a strong case against the defendant in any eventual jury trial. Such conditions will then cause juries to believe that defendants with strong evidence against them at trial are relatively likely to be guilty and therefore will be relatively likely to convict them. The second variable may affect the expected probability of conviction at trial if, conditional on any initial evidence against a defendant, any new evidence that arises in the time leading up to trial generally strengthens the case against a guilty defendant, but generally weakens the case against an innocent defendant.

Given the above argument, there are two primary ways in which plea bargaining may be used to manage society's desire to increase the expected punishment of the guilty and minimize the expected punishment of the innocent. First, related to point (i) above, if the likelihood of guilt is

increasing in the initial case against the defendant, prosecutors can increase the minimum plea bargain sentence they are willing to offer as the strength of the initial evidence against a defendant becomes stronger. This policy effectively increases the sentence length as the probability of guilt increases, subject to the plea bargain being acceptable to the defendant in lieu of a trial. Second, prosecutors may be able use plea bargaining as a type of screening mechanism to sort the guilty from the innocent. More specifically, related to point (ii) above, if conditional on the initial evidence against him, a guilty defendant believes he is more likely to be convicted at trial than an innocent defendant, a guilty defendant will be willing to accept a longer plea bargain sentence than an otherwise similar appearing innocent defendant. Thus, prosecutors may be able to use plea bargaining to get defendants to voluntarily reveal otherwise private information regarding their guilt.

At first glance, the above argument suggests that widespread plea bargaining may be an extremely effective tool for maximizing the expected punishment for those most likely to be guilty and minimizing the expected punishment of those most likely to be innocent. Specifically, by implementing a plea bargaining policy such that plea offers increase in severity with the strength of the initial evidence against a defendant, and conditional on the initial evidence against a defendant, the plea offer is sufficiently harsh such that only guilty defendants will accept it, prosecutors can ensure guilty individuals face an increasing punishment in the strength of the initial evidence against them. Moreover, given such a policy, essentially only innocent defendants choose to go to trial. This means that if juries are rational, they would be very unlikely to convict anyone at trial. Hence, such a policy may also help minimize the expected punishment of those who are innocent.

However, this last issue actually reveals the potential flaw in the above policy. Specifically, as alluded to above, what motivates guilty defendants to accept a plea deal is their expected likelihood of conviction by a jury at trial. Therefore, if a plea bargaining policy is implemented that screens most of the guilty away from going to trial, rational juries would be hesitant to convict anyone at trial, meaning any particular defendant (whether innocent or guilty) would only be willing to accept a very lenient plea bargain. This suggests plea bargaining may have a somewhat limited role when it comes to managing society's desire to severely punish the guilty but not punish the innocent. In particular, while high levels of plea bargaining can be used to

effectively sort the guilty from the innocent, the plea bargain sentences that can be imposed on the guilty under such a policy will have to be quite short. On the other hand, plea bargaining may be used to ensure that relatively long sentences are imposed on some guilty defendants, but only if such plea bargaining is limited to a relatively small fraction of defendants.

This article formalizes the intuitive argument discussed above. Namely, I consider an environment where only defendants know for sure whether they are guilty or innocent, jury behavior is determined endogenously along with prosecutor and defendant behavior, and prosecutors and juries each only observe a noisy signal regarding the underlying guilt of any given defendant. I make two key assumptions regarding these guilt signals. First, prosecutors observe an initial guilt signal at the time of arrest, where the likelihood that any given observation of this signal comes from a guilty individual *is increasing* as the signal increases. Second, juries are unable to observe the initial signal, but do observe a subsequent signal at trial, where this signal at trial is weakly greater than the defendant's initial signal at arrest if he is guilty, but is weakly less than the defendant's initial signal at arrest if he is innocent. Finally, I assume that defendants, prosecutors, and juries are rational, meaning their beliefs about unobservables must be consistent with the resulting truth in equilibrium.

I find that in this environment there actually exist a continuum of possible equilibria, where optimal prosecutor behavior differs across each of these equilibria, but can always be characterized in a similar fashion. Specifically, each equilibrium is characterized by a threshold value of the initial guilt signal such that for a defendant who emits an initial guilt signal below this characteristic threshold, it is optimal for the prosecutor to offer a plea bargain sentence that increases with the initial guilt signal, but is such that only a guilty defendant will accept it. In other words, in any equilibrium, for those defendants who emit an initial guilt signal below the relevant characteristic threshold, it is optimal for the prosecutor to use her discretion over plea offers to both tailor the sentence to the strength of the initial case against the defendant and also to use plea bargaining *to sort* the guilty from the innocent. However, for those defendants who emit an initial guilt signal greater than this relevant characteristic threshold, it is optimal for the prosecutor not to offer any plea bargain and instead take the case to trial.

What is most interesting, however, is how these equilibria differ across the continuum. Namely, there exists some lowest feasible equilibrium characteristic threshold where a relatively small fraction of guilty individuals resolve their cases through plea bargaining. In equilibria with characteristic thresholds near this lower bound, a relatively large fraction of those who go to trial are actually guilty, meaning juries will find it optimal to convict a defendant even if he emits a relatively low guilt signal at trial. This increases the probability of conviction at trial for any given initial guilt signal, increasing the plea bargain sentence each defendant is willing to accept. Therefore, in equilibria with low characteristic thresholds, relatively few cases are resolved through plea bargaining, meaning there is relatively little sorting, but those guilty defendants who do resolve their cases through plea bargains receive relatively long sentences.

On the other hand, in equilibria with higher characteristic thresholds, a larger fraction of guilty individuals resolve their cases through plea bargaining, meaning a smaller fraction of those who go to trial are guilty. This in turn means that in equilibria with relatively high characteristic thresholds, juries will find it optimal to convict a defendant only if he emits a relatively high guilt signal at trial, decreasing the probability of conviction at trial for any given initial guilt signal, thus lowering the plea bargain sentence a guilty defendant is willing to accept. This in turn means that in an equilibrium with a relatively high characteristic threshold, a relatively high fraction of cases are resolved through plea bargaining, meaning that plea bargaining may indeed be used to sort a relatively large fraction of the guilty from the innocent. However, in these equilibria that exhibit a high degree of sorting, it must be the case that those guilty defendants who resolve their cases through plea bargaining receive relatively short sentences.

Finally, I show that the range of feasible equilibria depends on the underlying efficiency of the policing and arrest process. Specifically, when the policing system is quite inefficient, it can be optimal for society to resolve a relatively high fraction of cases via plea bargaining even at the cost of high acquittal rates and short plea bargain sentences. However, as the policing system becomes more efficient, these equilibria where it is optimal to resolve a high fraction of cases via plea bargaining no longer remain feasible, meaning it is optimal for society to send a larger fraction of cases to trial, but impose relatively long sentences on those few who do plead guilty.

2. Related Literature

A variety of papers have theoretically examined the importance of accounting for imperfect information regarding defendant guilt in the criminal justice system (Andreoni, 1991; Friedman and Wickelgren, 2006; Gay et al., 1989; Miceli, 1990; Png, 1986; Rubinfield and Sappington, 1987). The papers most related to this one, however, are Grossman and Katz (1983), Reinganum (1988), and Baker and Mezzetti (2001). Like this article, these attempt to characterize how prosecutor discretion over plea bargaining offers can be used to manage society's desire to maximize the sentences imposed on guilty defendants while minimizing the sentences imposed on innocent defendants. In each case, the authors assume that both innocent and guilty individuals are arrested, but the probability that a jury convicts an innocent individual is strictly less than the probability that a jury convicts a guilty individual (although innocent individuals are still assumed to be convicted by juries with a strictly positive probability). Moreover, these authors assume that society incurs utility from punishing guilty individuals, but incurs disutility from punishing innocent individuals, and prosecutors at least partially act as agents of society.

Grossman and Katz (1983) first formalized the notion that plea bargaining may be used as a screening mechanism, sorting the guilty from the innocent. Intuitively, if guilty defendants believe they are more likely to be convicted if they go to trial than innocent defendants, they will be willing to accept longer plea bargain sentences than innocent defendants. Hence, plea bargaining can conceivably be used to make defendants voluntarily reveal otherwise unobservable information regarding their underlying guilt.

Reinganum (1988) provides an important innovation to Grossman and Katz's conception (1983) in that she assumes that not all defendants appear the same to prosecutors. Namely, prosecutors observe some information correlated with the guilt of the defendant prior to going to trial. In her model, prosecutors (but not the defendant, a judge, or a jury) observe a guilt signal at the time of arrest of each defendant, where this guilt signal represents the strength of the case against the defendant, with guilty individuals being more likely to have strong cases against them. Given such information, Reinganum shows that it may still be optimal for prosecutors to offer a plea bargain that only guilty defendants will accept, thereby using plea bargaining to sort the guilty from the innocent. However, Reinganum

also shows that given prosecutors observe this guilt signal prior to trial, under some conditions it is not optimal to use plea bargaining as a screening mechanism, but rather offer a plea bargain that all defendants are willing to accept, and tailor the offer to the strength of the case against the defendant, thereby increasing the sentences for those more likely to be guilty.

As motivated in the introduction, however, one drawback of these models is that when it comes to using plea bargaining as a sorting mechanism they implicitly assume that jury behavior is purely exogenous and cannot react to prosecutor behavior. If it could, sorting could not be maintained in equilibrium, as juries would never convict anyone if only innocent individuals went to trial, thereby erasing the incentive for guilty defendants to actually accept any plea offer.

Baker and Mezzetti (2001) raise this issue to some degree. In particular, while they still take jury behavior to be exogenous, they recognize that under perfect sorting where only innocent defendants reject plea bargains, prosecutors will not want to actually go ahead with any trials, as such trials could only lead to convicting innocent defendants. However, if prosecutors do not go ahead with trials, guilty defendants no longer have an incentive to accept a plea bargain. Hence, only a semi separating equilibrium can exist, with the plea offer being such that it is accepted by a fraction of the guilty defendants, with the remaining guilty defendants and all innocent defendants going to trial.

While Baker and Mezzetti's (2001) model brings up some of the logical difficulties inherent in using plea bargaining as a screening device, it does so by assuming that prosecutors cannot commit to go to trial if a plea bargain is not accepted, not by allowing jury behavior to be determined endogenously. Moreover, while Baker and Mezzetti's model allows for some heterogeneity among defendants, this heterogeneity is quite weak. Specifically, for each defendant, prosecutors are assumed to observe a signal that either reveals the defendant to be innocent, or reveals nothing, which means that all defendants who emit the innocent signal are released, and all the defendants who emit the uninformative signal appear identical to the prosecutor. Hence, among those defendants who are not immediately released, there is no heterogeneity. This means Baker and Mezzetti's model cannot be used to examine the optimality of prosecutors using their discretion to tailor the plea offer to the strength of the case against the defendant. As shown by Reinganum (1988), this may be important, as such

a strategy may be a preferred alternative to using plea bargaining as a screening mechanism.

The primary innovation of the model presented below is that it allows jury beliefs and behavior to be determined endogenously in equilibrium along with prosecutor and defendant behavior. Moreover, similar to Reinganum (1988), it also allows for substantial heterogeneity across defendants in terms of the evidence against them. However, I add two further features regarding this heterogeneity of evidence across defendants. First, consistent with rules of disclosure, I assume each defendant knows the strength of the case against him at the time of any plea bargaining negotiations, and second, I assume that juries may observe slightly different evidence against a defendant than was originally observed by the prosecutor at the time of arrest.

3. Model

Assume the law enforcement system is imperfect, meaning some fraction $\lambda \in (0, 1)$ of those who are arrested for any particular crime are guilty, while the rest are innocent. Let $j = G, I$ denote the two types of arrested defendants: those who are guilty and those who are innocent. While a defendant knows his type, the defendant's type is not directly observable to others. Rather, only imperfect signals of guilt can be observed, which can be interpreted as the strength of the evidence against the defendant. More specifically, at the time of arrest a defendant emits a guilt signal which is modeled as a random variable θ_a drawn from a cumulative distribution $F_j(\theta_a)$ defined over $\mathbb{R}$ given his type j. This initial guilt signal is observed by both the prosecutor and the defendant, as is consistent with the rules of disclosure. If we denote the pdf of $F_j(\theta_a)$ as $f_j(\theta_a)$, then the key assumption regarding the differences in the skill signal distributions across types will be that the likelihood ratio $\frac{f_I(\theta_a)}{f_G(\theta_a)}$ is strictly decreasing in θ_a and converges to zero as θ_a goes to infinity. Intuitively, the higher the observed initial guilt signal, the more likely it is to have come from a guilty individual, and moreover, the probability that a given signal comes from an innocent individual goes to zero as the signal becomes very large.

After an individual is arrested more evidence may arise in the time leading up to trial, meaning that if a defendant goes to trial, the jury may observe a different guilt signal than originally observed by the prosecutor.

In the case of an innocent defendant, any new evidence is likely to weaken the case against him, while for a guilty defendant any new evidence is likely to strengthen the case against him. To make this intuition explicit, assume that if an innocent defendant goes to trial, the jury will observe a guilt signal θ_t, where $\theta_t = \theta_a - \epsilon$, with ϵ being a random variable drawn from a distribution H, where $H(\epsilon) = 0$ for $\epsilon < 0$ and $H(\epsilon) > 0$ for $\epsilon \geq 0$. Alternatively, assume that if a guilty defendant goes to trial, the jury will observe a guilt signal θ_t, where $\theta_t = \theta_a + \epsilon$, and once again ϵ is a random variable drawn from the distribution H described above.

After observing the initial guilt signal θ_a for each defendant, the prosecutor must decide what plea bargain sentence to offer the defendant, which will be denoted as z_p. If the defendant accepts the plea bargain sentence, he receives a sentence of z_p and the judicial process ends for this defendant. If the defendant rejects this plea bargain, he goes to trial, where a jury observes the subsequent guilt signal θ_t and decides whether or not to convict.[1] If convicted by the jury, the defendant receives a sentence $\overline{z}$, where the prosecutor takes $\overline{z}$ as exogenously determined. If not convicted by the jury, the defendant receives no sentence.

Finally, assume individuals incur disutility of $-V(z)$ upon receiving a sentence length of z, where $V(0) = 0$, $V'(z) > 0$ and $V''(z) \geq 0$ for all $z > 0$ (i.e., individuals incur disutility from incarceration at an increasing rate, meaning defendants are risk averse).[2] Furthermore, assume that society's disutility from incarcerating an innocent individual is proportional to that individual's disutility. Specifically, if $-U_I(z)$ represents society's disutility associated with incarcerating an innocent individual for z months, assume $-U_I(z) = -\gamma V(z)$ for some parameter $\gamma > 0$. This means that the disutility the society incurs from increasing the sentence imposed on an innocent individual is also increasing at an increasing rate. This component of society's utility is depicted in Figure 1. Alternatively, assume society incurs disutility of $-U_G(z)$ for imposing a sentence of only length z on a guilty

1. Assume that the jury does not observe the initial guilt signal θ_a.

2. It is possible to assume that the disutility experienced by an innocent individual for any sentence z, denoted by $V_I(z)$, differs from that experienced by a guilty individual, denoted by $V_G(z)$. However, for the model's implications to stay unchanged, it must be assumed that $\frac{V_I(z)}{V_I(\overline{z})} \geq \frac{V_G(z)}{V_G(\overline{z})}$ for all $z \in (0, \overline{z}]$. Essentially, this assumes that innocent individuals are weakly more averse to shorter sentences than guilty individuals.

individual, where $-U_G(0) < 0$ and $U'_G(z) < 0$ and $U''_G(z) \geq 0$. In words, longer sentences imposed on guilty defendants increases society's utility, but at a weakly decreasing rate. This component of society's preferences is depicted in Figure 2.

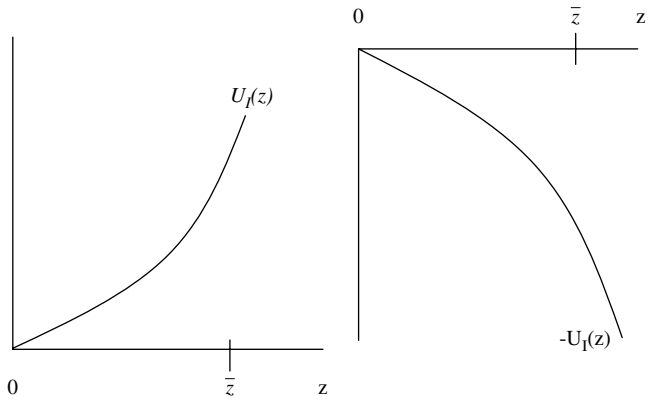

Figure 1. Underlying Shape for $U_I(z)$ Function and Depiction of $-U_I(z)$ (i.e., how societal utility decreases as the sentence length z imposed on an *innocent* defendant increases).

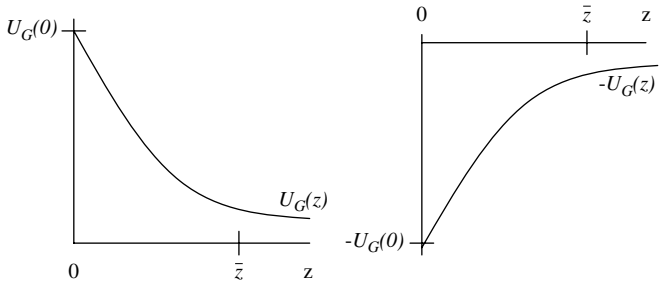

Figure 2. Underlying Shape for $U_G(z)$ Function and Depiction of $-U_G(z)$ (i.e., how societal utility increases as the sentence length z imposed on a *guilty* defendant increases).

Given the interest of this article lies in characterizing optimal prosecutor behavior with respect to plea bargaining as a means of managing society's

desire to increase punishment for the guilty and minimize punishment for the innocent, I assume prosecutors are acting as direct agents of society and trials are costless.[3] Therefore, the primary question asked in this article is, in the absence of trial costs, what plea bargain sentence should be offered to each defendant in equilibrium, if any, given each defendant's initial guilt signal θ_a? To answer this question, first I derive optimal defendant and prosecutor behavior given their beliefs about how jurors will behave. Then, I examine optimal jury behavior given their beliefs about how defendants and prosecutors behave. Finally, I examine what beliefs can be maintained by defendants, prosecutors, and juries in equilibrium if all agents are rational.

3.1. Defendant Behavior

Say defendants believe a jury will convict anyone who emits a guilt signal greater than some θ^* at trial. Give such beliefs, a defendant of type j will accept a plea bargain sentence length of z_p if and only if $-V(z_p) > -P(\theta_t \geq \theta^* | j, \theta_a) V(\overline{z})$, where $P(\theta_t \geq \theta^* | j, \theta_a)$ is the probability a type j defendant will emit a guilt signal at trial greater than θ^* given he emitted an initial guilt signal of θ_a. Rearranging and simplifying this expression given the assumptions above, we can see that a guilty defendant who emits an initial guilt signal of θ_a will accept a plea bargain of length z_p if and only if

$$1 - H(\theta^* - \theta_a) > \frac{V(z_p)}{V(\overline{z})}, \tag{1}$$

and an innocent defendant who emits an initial guilt signal of θ_a will accept a plea bargain of length z_p if and only if

$$H(\theta_a - \theta^*) > \frac{V(z_p)}{V(\overline{z})}. \tag{2}$$

As can be seen in Equation (1), given the distribution H equals zero when evaluated at any value less than zero, a guilty defendant will accept any plea bargain less than $\overline{z}$ if his initial θ_a exceeds θ^*. However, the smaller a guilty defendant's θ_a is compared to θ^*, the more of a discounted sentence the defendant will require in order to agree to a plea bargain. For notational

3. Therefore, in this environment, I abstract from my motivation for prosecutors to use plea bargaining as a way of managing scarce resources.

purposes, let $z_G(\theta_a)$ denote the maximum plea bargain sentence a guilty defendant who emits an initial guilt signal of θ_a is willing to accept. Given the discussion above, it should be clear that $z_G(\theta_a)$ is strictly increasing in θ_a up to the point where $\theta_a = \theta^*$, after which $z_G(\theta_a) = \overline{z} - \delta$ for $\theta_a > \theta^*$ (given any arbitrarily small $\delta > 0$).

Similarly, as can be seen in Equation (2), an innocent defendant will not accept any plea bargain if his initial θ_a is less than θ^*. However, the larger an innocent defendant's θ_a is compared to θ^*, the longer the plea bargain such an innocent defendant will be willing to accept, subject to this sentence being short of $\overline{z}$. Therefore, if we let $z_I(\theta_a)$ denote the maximum plea bargain sentence an innocent defendant who emits an initial guilt signal of θ_a is willing to accept, it will be the case that $z_I(\theta_a)$ will equal zero for all $\theta_a \leq \theta^*$, but will be strictly increasing in θ_a and less than $\overline{z}$ for all $\theta_a > \theta^*$.

A direct implication of the above argument is that $z_G(\theta_a)$ will be greater than $z_I(\theta_a)$ for all θ_a. Intuitively, for any given strength of pre-trial evidence, guilty individuals will be willing to accept longer plea bargain sentences than innocent individuals, as innocent individuals will believe that any further evidence that arises prior to their case will likely help their case, while just the opposite will be true for guilty individuals.

Finally, as will become important in equilibrium, note that both $z_G(\theta_a)$ and $z_I(\theta_a)$ are decreasing in θ^* for all θ_a. Intuitively, if defendants believe juries will require a very high standard of evidence in order to convict a defendant at trial (i.e., a relatively high guilt signal threshold at trial), defendants will require a substantially discounted sentence in order to be willing to accept a plea bargain.

3.2. Prosecutor Behavior

The defendant behavior derived above suggests three relevant strategies for prosecutors with respect to the plea bargain sentence to offer a defendant with an initial guilt signal θ_a. If the prosecutor offers a plea bargain sentence length $z_p \leq z_I(\theta_a)$, the defendant will accept the plea bargain regardless of whether he is guilty or innocent. If the prosecutor offers a plea bargain sentence length $z_I(\theta_a) < z_p \leq z_G(\theta_a)$, the defendant will accept the plea bargain if he is guilty and reject the plea bargain if he is innocent. Finally, if the prosecutor offers $z_p > z_G(\theta_a)$ the defendant will choose to go to trial whether he is guilty or innocent. In other words, for any given initial guilt signal θ_a, the prosecutor may choose to send all such cases to trial, use plea

bargaining to sort the guilty from the innocent, or use plea bargaining to avoid the uncertainty of trial for all defendants.

Before deriving optimal prosecutor behavior, assume that prosecutors also believe that juries will convict any defendant who emits a guilt signal θ_t greater than some θ^* at trial. Given such beliefs, there are two things to note. First, under the sorting strategy where the prosecutor offers a plea sentence $z_I(\theta_a) < z_p \leq z_G(\theta_a)$, the prosecutor's expected utility will equal

$$E[U(z_p)|z_I(\theta_a) < z_p \leq z_G(\theta_a), \theta_a] =$$
$$-P(G|\theta_a)U_G(z_p) - [1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z}),$$

where $P(G|\theta_a)$ is the probability a defendant is guilty given an initial guilt signal θ_a, and $P(\theta_t \geq \theta^*|I, \theta_a)$ is the probability a defendant will emit a guilt signal greater than θ^* given he is innocent (i.e. type I) and emitted an initial guilt signal of θ_a. Since z_p appears only in the left term (since innocent individuals will not accept the plea bargain in this range) and $U'_G < 0$, it will be true that $E[U(z_p)|z_I(\theta_a) < z_p \leq z_G(\theta_a), \theta_a]$ is strictly increasing in z_p for all θ_a. *Therefore, if the prosecutor wants to offer a plea bargain in the sorting range, it is optimal to set it to the upper bound of this range, or equal to* $z_G(\theta_a)$. Intuitively, since all innocent defendants will choose to go to trial for plea offers in this range, the prosecutor wants to maximize the sentence offered in the plea bargain since he knows anyone who accepts the offer is guilty. This means that if the prosecutor is behaving optimally and decides to offer a plea bargain that sorts the guilty from the innocent, then her utility will equal

$$E[U(z_p)|z_p = z_G(\theta_a), \theta_a] =$$
$$-P(G|\theta_a)U_G(z_G(\theta_a)) - [1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z}). \quad (3)$$

The second thing to note before proceeding is that if the prosecutor offers a plea bargain sentence z_p greater than $z_G(s)$, no defendant will accept it, preferring to take his chances at trial instead. Therefore, for all plea offers such that $z_p > z_G(\theta_a)$, the prosecutor is essentially choosing to

send all cases to trial, meaning her expected utility will equal

$$E[U(z)|z_p > z_G(\theta_a), \theta_a] = -[1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z})$$
$$-P(G|s)[P(\theta_t \geq \theta^*|G, \theta_a)U_G(\overline{z}) + (1 - P(\theta_t \geq \theta^*|G, \theta_a))U_G(0)]. \quad (4)$$

Given the above equations, we can now state and prove the following Lemma.

Lemma 1. *Given defendants and prosecutors believe juries will convict all defendants who emit a guilt signal at trial θ_t greater than some threshold level θ^*, it is never optimal for the prosecutor to offer a plea sentence z_p^* where $z_p^* \leq z_I(\theta_a) < z_G(\theta_a)$, such that both innocent and guilty defendants find it optimal to accept it.*

Proof. As discussed above, one strategy for the prosecutor is to offer a plea bargain sentence z_p such that both innocent and guilty defendants accept it. By offering a plea sentence equal to some z_p^* where $z_p^* \leq z_I(\theta_a) < z_G(\theta_a)$, the prosecutor's expected utility will equal

$$E[U(z_p)|z_p = z_p^* \leq z_I(\theta_a) < z_G(\theta_a), \theta_a] =$$
$$-P(G|\theta_a)U_G(z_p^*) - [1 - P(G|\theta_a)]U_I(z_p^*). \quad (5)$$

Let us first consider whether it is ever optimal to offer such a sentence to a defendant who emits an initial guilt signal $\theta_a \leq \theta^*$. Equation (3) gives the prosecutor's expected utility from offering a defendant a plea offer equal to $z_G(\theta_a)$. However, recall that since it was assumed that the strength of case against innocent defendants would not become stronger between the time of initial arrest and trial, for defendants who emit an initial guilt signal $\theta_a \leq \theta^*$, it will be the case that $P(\theta_t \geq \theta^*|I, \theta_a \leq \theta^*) = 0$, meaning Equation (3) reduces to

$$E[U(z_p)|z_p = z_G(\theta_a), \theta_a \leq \theta^*] = -P(G|\theta_a)U_G(z_G(\theta_a)) \quad (6)$$

for such defendants. Recalling that $U_G(z)$ is decreasing in z, it is straightforward from Equations (5) and (6) to see that for defendants who emit an initial guilt signal $\theta_a \leq \theta^*$ prosecutors have higher expected utility from offering $z_G(\theta_a)$ to any plea sentence z_p^* where $z_p^* \leq z_I(\theta_a) < z_G(\theta_a)$.

Now let us consider whether it is ever optimal to offer a plea sentence z_p^*, where $z_p^* \leq z_I(\theta_a) < z_G(\theta_a)$, to a defendant who emits an initial guilt signal $\theta_a > \theta^*$. In this case, recall that since it was assumed that the strength of case against guilty defendants would never become weaker between the time of initial arrest and trial, for defendants who emit an initial guilt signal $\theta_a > \theta^*$, it will be the case that $P(\theta_t \geq \theta^* | G, \theta_a > \theta^*) = 1$. Applying this result to Equation (4), we see that the prosecutor's expected utility from offering a plea sentence $z_p > z_G(\theta_a)$ to a defendant who emits an initial guilt signal $\theta_a > \theta^*$ will equal

$$E[U(z)|z_p > z_G(\theta_a), \theta_a > \theta^*] =$$
$$-P(G|\theta_a)U_G(\overline{z}) - [1 - P(G|\theta_a)]P(\theta_t \geq \theta^* | I, \theta_a > \theta^*)U_I(\overline{z}). \quad (7)$$

Comparing Equations (5) to (7), we can see that a sufficient condition for the prosecutor to be worse off by offering a plea sentence less than $z_I(\theta_a)$ to a defendant who emitted an initial guilt signal $\theta_a > \theta^*$ is for $-[1 - P(G|\theta_a)]U_I(z_p^*) < -[1 - P(G|\theta_a)]P(\theta_t \geq \theta^* | I, \theta_a)U_I(\overline{z})$ for all $z_p^* < z_I(\theta_a)$, or equivalently when

$$\frac{U_I(z_p^*)}{U_I(\overline{z})} > P(\theta_t \geq \theta^* | I, \theta_a > \theta^*).$$

Recalling that $-U_I(z) = -\gamma V(z)$ for some parameter $\gamma > 0$ and the other assumptions made previously, the above condition reduces to

$$\frac{V(z_p^*)}{V(\overline{z})} > H(\theta_a - \theta^*). \quad (8)$$

Now, recalling from the discussion of optimal defendant behavior that for all $z_p \leq z_I(\theta_a)$ and $\theta_a > \theta^*$ we know $\frac{V(z_p)}{V(\overline{z})} \leq H(\theta_a - \theta^*)$, meaning Equation (8) can never hold for any $z_p^* < z_I(\theta_a)$. Therefore, it is also the case that for defendants who emit an initial guilt signal $\theta_a > \theta^*$, it again can never be optimal for the prosecutor to offer a plea sentence z_p^* where $z_p^* \leq z_I(\theta_a) < z_G(\theta_a)$. ■

Given that it is never optimal for a prosecutor to offer a plea sentence that both guilty and innocent defendants would accept, deriving optimal prosecutor behavior (given beliefs about jury behavior) reduces to determining whether there exist defendants for which it is optimal to use plea

bargaining to sort the guilty from the innocent or whether all defendants should simply be sent to trial. This leads to Theorem 1.

Theorem 1. *Given defendants and prosecutors believe juries will convict all defendants who emit a guilt signal at trial θ_t greater than some threshold level θ^*, then:*

1. For defendants who emit initial guilt signals below this threshold ($\theta_a \leq \theta^*$) it is optimal for the prosecutor to use plea bargaining to sort the innocent from the guilty by offering a plea sentence given by the function $z_G(\theta_a)$, which is increasing in the initial guilt signal θ_a.
2. For defendants who emit initial guilt signals greater than this threshold ($\theta_a > \theta^*$) it is optimal for the prosecutor not to offer a plea sentence that is acceptable to either guilty or innocent defendants.

Proof. In Appendix. ■

3.3. Jury Behavior

Given that juries are assumed to be made up of members of society, let us assume they have the same preferences as society. Therefore, given a defendant deemed guilty by a jury will receive a sentence length of $\overline{z}$, the jury's expected utility for convicting a defendant who emits a guilt signal at trial of θ_t is $-\pi(G|\theta_t) - [1 - \pi(G|\theta_t)]U_I(\overline{z})$, where $\pi(G|\theta_t)$ is the jury's belief regarding the probability that a defendant who emits a guilt signal of θ_t is actually guilty. Alternatively, the jury's expected utility for not convicting a defendant who emits a guilt signal of θ_t at trial is $-U_G(0)\pi(G|\theta_t)$. Therefore, a jury will convict an individual if and only if he emits a guilt signal at trial such that $-\pi(G|\theta_t)U_G(\overline{z}) - [1 - \pi(G|\theta_t)]U_I(\overline{z}) \geq -U_G(0)\pi(G|\theta_t)$, or if and only if a defendant emits a guilt signal at trial θ_t such that the following expression holds:

$$\frac{\pi(G|\theta_t)}{1 - \pi(G|\theta_t)} \geq \frac{U_I(\overline{z})}{U_G(0) - U_G(\overline{z})}. \tag{9}$$

3.4. Feasible Equilibria

The relevant equilibrium in this context is a *Perfect Bayesian Equilibrium* (PBE), which in this environment is defined to be a set of behavioral rules and beliefs for each actor in the judicial system such that these behavioral rules dictate optimal behavior given each actor's beliefs, and

each actor's beliefs are consistent with Bayes' rule given all other actors behave in accordance with these optimal rules. Intuitively, a PBE is a Nash Equilibrium given each participants' beliefs, with the condition that these beliefs conform to the true state of the world in equilibrium. This section attempts to characterize the existence of such equilibria in this environment.

The above analysis showed optimal behavior for each actor to be the following. First, assuming defendants believe that juries will convict any defendant who emits a guilt signal greater than some θ^* at trial, it is optimal for a defendant who emits an initial guilt signal of θ_a to only accept a plea bargain sentence less than or equal to $z_G(\theta_a)$ (as given implicitly by Equation (1)) if he is guilty, and less than or equal to $z_I(\theta_a)$ (as given implicitly by Equation (2)) if he is innocent, where both $z_G(\theta_a)$ and $z_I(\theta_a)$ are weakly increasing in θ_a and $z_G(\theta_a) > z_I(\theta_a)$ for all θ_a. Second, assuming prosecutors believe that juries will convict any defendant who emits a guilt signal greater than some θ^* at trial, it is optimal for prosecutors to offer a plea bargain sentence of length $z_G(\theta_a)$ to all defendants who emit an initial guilt signal θ_a less than or equal to θ^*, and send all defendants to trial who emit initial guilt signals greater than θ^*. Finally, given a jury believes that the probability of guilt for a defendant who emits a guilt signal at trial of θ_t equals $\pi(G|\theta_t)$, it is optimal to convict a defendant if and only if he emits a guilt signal at trial θ_t such that Equation (9) holds.

The key issue that remains is to determine what defendant and prosecutor beliefs concerning jury behavior can be maintained in a PBE, and what this implies about what can constitute optimal prosecutor behavior. As discussed in the earlier section regarding the previous literature in this area, if prosecutors use plea bargaining to perfectly sort the guilty from the innocent, rational juries will never convict anyone who goes to trial (given they are all innocent), meaning a rational defendant would turn down all plea offers and go to trial regardless of his underlying guilt, no longer making jury behavior optimal. Hence, with rational judicial agents, it is not possible to achieve perfect sorting in equilibrium. However, in the environment discussed above, it was shown that if prosecutors believe juries will convict a defendant if and only if he emits a guilt signal at trial greater than some threshold θ^*, optimal prosecutor behavior with respect to plea bargaining would not exhibit perfect sorting. Rather, prosecutors only use plea bargaining to sort the innocent from the guilty for those with initial guilt signals under this threshold, with all defendants whose initial guilt

signals exceed this threshold going to trial. The remainder of this section discusses which, if any, such thresholds can be maintained in equilibrium.

To start this discussion, first note that given prosecutors offer plea bargains such that individuals choose to go to trial only if they are innocent or if they are guilty and emitted an initial guilt signal greater than some θ^*, then according to Bayes' rule, equilibrium jury beliefs regarding the probability of guilt for a defendant who emits a guilt signal at trial of θ_t must be such that

$$\pi(G|\theta_t) = \begin{cases} 0 & \text{if } \theta_t \leq \theta^* \\ \dfrac{\lambda \int_{\theta^*}^{\theta_t} f_G(x)h(\theta_t - x)dx}{\lambda \int_{\theta^*}^{\theta_t} f_G(x)h(\theta_t - x)dx + (1-\lambda)\int_{\theta_t}^{\infty} f_I(x)h(x-\theta_t)dx} & \text{if } \theta_t > \theta^* \end{cases}. \quad (10)$$

To determine the existence of an equilibrium, we must then determine whether there exist any possible values for θ^* such that when plugged into Equation (10) the resulting decision rule for juries leads to jury outcomes consistent with the assumed defendant and prosecutor beliefs (i.e., that juries convict any defendant who emits a guilt signal at trial θ_t that is greater than this θ^*). To put this in another way, do there exist any possible values for θ^* such that if prosecutors offer plea bargains only to defendants who emit an initial guilt signal less than this θ^*, and only guilty defendants accept such offers, then juries find it optimal to convict only defendants who emit a guilt signal at trial that is greater than this θ^* given their beliefs are consistent with Equation (10)?

To examine this question, we can substitute Equation (10) into Equation (9). Doing so, it is straightforward to see that, for any given θ^* a jury will indeed find it optimal not to convict any defendant who emits a guilt signal at trial less than or equal to this θ^*. What remains to be determined, however, is whether there exist any possible values for θ^* such that given the optimal defendant and prosecutor behavior that result from this θ^*, juries also find it optimal to convict all defendants who emit a guilt signal at trial that is greater than this θ^*. To examine this, we note that for $\theta_t > \theta^*$, when we substitute Equation (10) into Equation (9) and rearrange, we see that an equilibrium exists if and only if there exist any θ^* such that the expression

$$\frac{\lambda}{1-\lambda}\frac{U_G(0) - U_G(\overline{z})}{U_I(\overline{z})} \geq \frac{\int_{\theta_t}^{\infty} f_I(x)h(x-\theta_t)dx}{\int_{\theta^*}^{\theta_t} f_G(x)h(\theta_t - x)dx} \quad (11)$$

holds for all $\theta_t > \theta^*$. Clearly, for such a θ^* to exist, it must be the case that $f_I(\theta_a)$ becomes very small relative to $f_G(\theta_a)$ as θ_a gets large, which will be true given the assumption that the likelihood ratio $\frac{f_I(\theta_a)}{f_G(\theta_a)}$ goes to zero as θ_a goes to infinity. Moreover, given the decreasing likelihood ratio assumption, if Equation (11) holds for some θ_1^* then it also holds for all $\theta^* > \theta_1^*$. Hence, there will be a continuum equilibria, each characterized by a different threshold value θ^*.

Like in the Baker and Mezzetti (2001) model, where jury behavior was taken as exogenous but prosecutors could not commit to taking a case to trial, the equilibria in this environment will be semiseperating, with plea bargains only being taken by a fraction of the guilty and the remaining fraction of the guilty as well as all of the innocent choosing to go to trial. However, unlike in Baker and Mezzetti's model, in this model where jury behavior is endogenous, there is a multiplicity of equilibria where each differs in the degree of separation that arises. Specifically, in equilibria with a very high θ^*, optimal plea bargaining will indeed lead to almost complete sorting. What is also interesting, however, is that a high θ^* will mean $z_G(\theta_a)$ will be very small for all θ_a, as can be confirmed by looking at Equation (1) (i.e., the equation that summarizes optimal behavior for guilty defendants). Intuitively, for a θ^* that is very high, defendants of either type will be very unlikely to emit a guilt signal at trial that juries would deem sufficient for conviction, meaning to get a guilty defendant to accept a plea deal, the prosecutor would have to offer a very short sentence. On the other hand, for a θ^* that is relatively low, optimal plea bargaining will lead to much less sorting between the guilty and innocent, but the prosecutor will be able to impose much longer sentences on those defendants who do plead guilty.

This multiplicity of equilibria reveals the importance of endogenizing jury beliefs. Specifically, beliefs can be self-reinforcing. If prosecutors believe juries will be relatively likely to convict any given defendant, it is optimal for prosecutors to offer plea bargains (that are acceptable only to guilty defendants) to only a relatively small fraction of defendants. This means a relatively large fraction of those going to trial are guilty, making it optimal for juries to set a relatively low standard of evidence required for conviction, thus confirming beliefs that convictions at trial are relatively likely. Alternatively, if prosecutors believe juries will be relatively unlikely to convict any given defendant, it is optimal for prosecutors to offer plea bargains (that are acceptable only to guilty defendants) to a relatively large

fraction of defendants. This in turn means that a relatively large fraction of those who go to trial are innocent, making it optimal for juries to set a relatively high standard of evidence required for conviction at trial, thus confirming beliefs that convictions at trial are relatively unlikely.

These results also reveal the theoretical limitations of using plea bargaining as a method for managing society's desire to impose long sentences on the guilty, but minimize the sentences imposed on the innocent. Specifically, extensive plea bargaining can theoretically induce a large fraction of the guilty to sort themselves from the innocent in equilibrium, but only at the cost of imposing relatively short sentences on those who do so. Alternatively, plea bargaining can be used to ensure that some guilty defendants receive relatively long sentences, but only if such plea bargaining is used in a relatively limited fashion.

A further thing to note from Equation (11) is with respect to the lowest feasible equilibrium threshold, or the lowest θ^* for which Equation (11) holds, which we can denote as θ^*_ℓ. Specifically, for any given set of societal preferences, if a high fraction of those arrested are guilty (i.e., λ is large), then θ^*_ℓ will be relatively low. On the other hand, when a sizable fraction of those arrested are actually innocent (i.e., $1 - \lambda$ is large), θ^*_ℓ will be relatively high. Hence, this analysis formally shows that when the policing system is quite efficient, it will generally be optimal for society to send a very large fraction of cases to trial, meaning very little sorting will take place, but relatively long sentences can be imposed on the relatively few guilty individuals who accept plea bargains. Alternatively, when the policing system is quite inefficient, it can actually be optimal for society to resolve a relatively high fraction of cases via plea bargaining even at the cost of high acquittal rates and relatively short sentences being imposed on those who accept plea bargains. Interestingly, this result is surprisingly similar to a key result coming from Reinganum (1988) where jury behavior was assumed to be exogenous and did not depend on the degree of sorting used by the prosecutor.

Finally, these results were all derived under the assumption that prosecutor decisionmaking is not affected by the additional costs associated with a trial. If prosecutors did weigh the cost of a trial into their decisionmaking, then it would be even more difficult for prosecutors to use plea bargaining as a sorting device. Intuitively, the only way plea bargaining can be used as a sorting device with rational juries is if such sorting is only partial, meaning some guilty defendants must go to trial. As shown above, without

trial costs prosecutors find it optimal to send all of those defendants who emit a sufficiently high initial guilt signal, to trial (i.e., those with $\theta_a > \theta^*$). This assures some guilty individuals going to trial. However, with trial costs the prosecutor would be willing to offer a bit of a discounted sentence to those defendants with these high initial guilt signals, possibly causing all of the guilty defendants with high initial guilt signals to also accept a plea. This would further undermine the existence of any equilibrium where plea bargaining sorts the guilty from the innocent.

4. Summary and Conclusion

This article examined optimal prosecutor behavior with respect to plea bargaining when defendant guilt is uncertain, defendants differ in the strength of the initial evidence against them, juries observe slightly different evidence against each defendant who goes to trial than was originally observed by the prosecutor, and in equilibrium, defendant, prosecutor, jury beliefs are rational given the behavior of others. I show that in this environment, there exist a continuum of possible equilibria, where optimal prosecutor behavior is similar in character but varies in important ways across equilibria. Specifically, in all equilibria, it is optimal for prosecutors to offer plea bargains to only a fraction of defendants, where the offered plea bargain sentence increases in the strength of the initial evidence against the defendant, and is such that only a guilty defendant would find it acceptable.

What differs across equilibria, however, is the fraction of defendants who are offered plea bargains, the probability of being convicted at trial for any given defendant, and the length of the plea bargain sentence defendants find acceptable. Namely, in an equilibrium where prosecutors use a plea bargaining strategy such that primarily only innocent defendants choose to go to trial rather than accept a plea bargain, juries are rationally quite hesitant to convict anyone without observing extremely strong evidence against them. This means defendants would be unwilling to accept a plea bargain unless it offers a substantial discount relative to the sentence they would receive if convicted at trial. On the other hand, if prosecutors offer potentially acceptable plea bargains to only the relatively small fraction of defendants with weak initial cases against them, a substantial fraction of those going to trial will be guilty, meaning rational juries will find it optimal to convict even those with relatively modest evidence against them. This

relatively greater certainty of conviction at trial will in turn mean that those defendants who do receive plea offers are willing to accept relatively long plea bargain sentences.

These results reveal that when jury behavior is determined in equilibrium rather than assumed to be exogenous, *plea bargaining can generally play only a limited role when it comes to managing society's desire to impose severe punishments on the guilty but not punish the innocent*. While it is admittedly a strong assumption that equilibrium jury beliefs correspond identically to the true probability that any given defendant is guilty given the observed evidence, this assumption can certainly be interpreted as an approximation, where this approximation ensures that wildly irrational beliefs do not drive the results. Therefore, this analysis suggests that while plea bargaining may have a variety of theoretical justifications, it is questionable whether it can be justified on the grounds that it can act as an effective tool for managing society's conflicting desires regarding punishment.

Appendix - Proof of Theorem 1

Given Lemma 1, to derive optimal behavior it is sufficient to see if there are any initial guilt signals θ_a such that Equation (3) (i.e., the prosecutor's expected utility from offering the optimal sorting plea bargain) exceeds Equation (4) (i.e., the prosecutor's expected utility from taking the case to trial). To examine this, it is once again helpful to look first at defendants who emit initial guilt signals less than or equal to θ^*, and then look at optimal prosecutor behavior with respect to defendants who emit initial guilt signals greater than θ^*.

As discussed above, for any defendant who emits an initial guilt signal $\theta_a \leq \theta^*$, it will be the case that by offering a plea sentence equal to $z_G(\theta_a)$, a prosecutor incurs an expected utility of

$$E[U(z_p)|z_p = z_G(\theta_a), \theta_a \leq \theta^*] =$$
$$-P(G|\theta_a)U_G(z_G(\theta_a)) - [1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z}).$$

Similarly, given Equation (4), a prosecutor's expected utility from taking a defendant who emits an initial guilt signal $\theta_a \leq \theta^*$ to trial equals

$$E[U(z)|z_p > z_G(\theta_a), \theta_a \leq \theta^*] =$$
$$-P(G|\theta_a)[P(\theta_t \geq \theta^*|G, \theta_a)U_G(\overline{z}) + (1 - P(\theta_t \geq \theta^*|G, \theta_a))U_G(0)]$$
$$-[1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z})$$

Using the fact that $P(\theta_t \geq \theta^* | I, \theta_a) = 0$ when $\theta_a \leq \theta^*$, the two equations above imply that it is only optimal for the prosecutor to offer a plea sentence equal to $z_G(\theta_a)$ to a defendant who emits an initial guilt signal equal to $\theta_a \leq \theta^*$ if $U_G(z_G(\theta_a)) \leq P(\theta_t \geq \theta^* | G, \theta_a \leq \theta^*) U_G(\overline{z}) + (1 - P(\theta_t \geq \theta^* | G, \theta_a \leq \theta^*)) U_G(0)$. Re-writing this expression after recalling that $P(\theta_t \geq \theta^* | G, \theta_a) = 1 - H(\theta^* - \theta_a)$ and $1 - H(\theta^* - \theta_a) = \frac{V(z_G(\theta_a))}{V(\overline{z})}$ for all $\theta_a \leq \theta^*$, we can see that it is optimal for the prosecutor to offer $z_G(\theta_a)$ to a defendant who emits an initial guilt signal equal to $\theta_a \leq \theta^*$only if

$$\frac{V(z_G(\theta_a))}{V(\overline{z})} \leq \frac{U_G(0) - U_G(z_G(\theta_a))}{U_G(0) - U_G(\overline{z})}.$$

To show that the above expression holds for all $z_G(\theta_a) < \overline{z}$, let us define $\widetilde{U}_G(z) = U_G(0) - U_G(z)$, making the above condition equivalent to

$$\frac{V(z_G(\theta_a))}{V(\overline{z})} \leq \frac{\widetilde{U}_G(z_G(\theta_a))}{\widetilde{U}_G(\overline{z})}. \tag{12}$$

Given $U'_G(z) < 0$ and $U''_G(z) \leq 0$, it is straightforward to confirm that $\widetilde{U}'_G(z) > 0$ and $\widetilde{U}''_G(z) \leq 0$. Moreover, it was assumed at the outset that $V'(z) > 0$ and $V''(z) \geq 0$. In other words, $\widetilde{U}_G(z)$ is an increasing concave function and $V(z)$ is an increasing convex function. Given this, we know $\widetilde{U}_G(\overline{z}) - \widetilde{U}_G(z) \leq \overline{z} - z \leq V(\overline{z}) - V(z)$ for any given $z < \overline{z}$. Therefore, it is also true that $\widetilde{U}_G(z) - \widetilde{U}_G(\overline{z}) \geq V(z) - V(\overline{z})$. Since the natural logarithm is an increasing monotonic transformation, it is also true that $\ln(\widetilde{U}_G(z) - \widetilde{U}_G(\overline{z})) \geq \ln(V(z) - V(\overline{z}))$, and equivalently that $\ln \frac{\widetilde{U}_G(z)}{\widetilde{U}_G(\overline{z}))} \geq \ln \frac{V(z)}{V(\overline{z})}$. Once again recognizing that the natural log is a strictly increasing function, the previous equation implies that $\frac{\widetilde{U}_G(z)}{\widetilde{U}_G(\overline{z}))} \geq \frac{V(z)}{V(\overline{z})}$ for all $z < \overline{z}$, confirming that Equation (12) holds for all $z_G(\theta_a) < \overline{z}$, and hence, it is optimal for the prosecutor to offer $z_G(\theta_a)$ to every defendant who emits an initial guilt signal less than or equal to θ^*.

Now consider optimal prosecutor behavior with respect to those defendants who emit initial guilt signals greater than θ^*. First, noting Lemma 1 and recalling that when offering a plea greater than $z_I(\theta_a)$ it is never optimal for a prosecutor to offer a plea less than $z_G(\theta_a)$, we know that the prosecutor will never offer a plea sentence less than $z_G(\theta_a)$. Moreover, given that $z_G(\theta_a) = \overline{z} - \delta$ for all $\theta_a \geq \theta^*$ (as argued in the text) given any arbitrarily small $\delta > 0$ then the prosecutor's expected utility from plea bargaining equals

$$E[U(z_p) | z_p = z_G(\theta_a), \theta_a > \theta^*] =$$
$$-P(G|\theta_a) U_G(z_G(\theta_a) - \delta) - [1 - P(G|\theta_a)] P(\theta_t \geq \theta^* | I, \theta_a) U_I(\overline{z}).$$

Alternatively, by taking such a defendant to trial, the prosecutor's expected utility will equal

$$E[U(z)|z_p > z_G(\theta_a), \theta_a > \theta^*] =$$
$$-P(G|\theta_a)[P(\theta_t \geq \theta^*|G, \theta_a)U_G(\overline{z}) + (1 - P(\theta_t \geq \theta^*|G, \theta_a))U_G(0)]$$
$$-[1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z})$$

Recalling that $P(\theta_t \geq \theta^*|G, \theta_a) = 1$ when $\theta_a > \theta^*$, the above equations simplify to

$$E[U(z_p)|z_p = z_G(\theta_a), \theta_a > \theta^*] = -P(G|\theta_a)U_G(\overline{z} - \delta)$$
$$-[1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z})$$

and

$$E[U(z)|z_p > z_G(\theta_a), \theta_a > \theta^*] = -P(G|\theta_a)U_G(\overline{z})$$
$$-[1 - P(G|\theta_a)]P(\theta_t \geq \theta^*|I, \theta_a)U_I(\overline{z}).$$

Comparing these two expressions, it is clear that $E[U(z_p)|z_p = z_G(\theta_a), \theta_a > \theta^*] < E[U(z)|z_p > z_G(\theta_a), \theta_a > \theta^*]$, meaning the prosecutor should take all defendants who emit a $\theta_a > \theta^*$ signal to trial, confirming Part 2 of Theorem 1.

References

Andreoni, James. 1991. "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty fit the Crime?" 22 *RAND Journal of Economics* 385–95.

Baker, Scott, and Claudio Mezzetti. 2001. "Prosecutorial Resources, Plea Bargaining, and the Decision to Go to Trial," 17 *Journal of Law, Economics, and Organization* 149–67.

Friedman, Ezra, and Abraham L. Wickelgren. 2006. "Bayesian Jurors and the Limits to Deterrence," 22 *Journal of Law, Economics, and Organization* 70–86.

Gay, Gerald D., Martin F. Grace, Jayant R. Kale, and Thomas H. Noe. 1989. "Noisy Juries and the Choice of Trial Mode in a Sequential Signalling Game: Theory and Evidence," 20 *RAND Journal of Economics* 196–213.

Grossman, Gene M., and Michael L. Katz. 1983. "Plea Bargaining and Social Welfare," 73 *American Economic Review* 749–57.

Miceli, Thomas J. 1990. "Optimal Prosecution of Defendants Whose Guilt is Uncertain," 6 *Journal of Law, Economics, and Organization* 189–201.

Png, I. P. L. 1986. "Optimal Subsidies and Damages in the Presence of Judicial Error," 6 *International Review of Law and Economics* 101–5.

Reinganum, Jennifer F. 1988. "Plea Bargaining and Prosecutorial Discretion," 78 *American Economic Review* 713–28.

Rubinfeld, Daniel L., and David E. M. Sappington. 1987. "Efficient Awards and Standards of Proof in Judicial Proceedings," 18 *RAND Journal of Economics* 308–15.