DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

STRATEGIC VOTING IN A JURY TRIAL WITH PLEA BARGAINING

SangMok Lee

SOCIAL SCIENCE WORKING PAPER 1327

June 2010

Strategic Voting in a Jury Trial with Plea Bargaining

SangMok Lee

Abstract

We study the criminal court process focusing on the interaction between plea bargaining and jury trials. We model plea bargaining such that a prosecutor makes a take-it-or-leave-it offer and a defendant, who is either guilty or innocent, pleads either guilty or not guilty. If the defendant pleads not guilty, the case goes to a jury trial, which follows a strategic voting model. Plea bargaining produces a bias in which the defendant is less likely to be guilty if the case goes to trial, which in turn alters the jurors' voting behavior. Conversely, anticipated jury trial outcomes affect a prosecutor and a defendant while they participate in a plea bargain. We find that the equilibrium behavior in a court with plea bargaining and a jury trial, resembles the equilibrium behavior in the separate jury model, though jurors may act as if they echo the prosecutor's preference against convicting the innocent and acquitting the guilty. We also compare two voting paradigms, unanimity and non-unanimity. The unanimity rule is inferior to non-unanimity because the ex-ante punishment delivered to the innocent or undelivered to the guilty by unanimity rule does not vanish as the size of jury gets large.

JEL classification numbers: D72, K41

Key words: Plea Bargaining, Strategic Voting, Condorcet Jury Theorem, Criminal Court

Strategic Voting in a Jury Trial with Plea Bargaining

SangMok Lee *

1 Overview

1.1 Introduction

Plea bargaining is a pre-trial stage where a defendant is allowed to plead guilty. A defendant pleads guilty primarily in exchange for a lesser charge than he would receive if he was convicted after a jury trial.[1] Plea bargaining is so prevalent that amongst the 89.7% convicted out of 83,391 defendants in Federal Courts in 2004, 96% of them were by pleading guilty. For felony offenses, 96% of convicted defendants pleaded guilty, which is an increase from 87% in 1990.[2] The fact that the vast majority of cases end in plea bargaining may cause many people to believe that trials are not important.

However, such a conclusion is inaccurate; plea bargaining and jury trials closely interact with each other, and the interaction plays an important role in the judicial process. Although most cases are settled before jury trials begin, participants in plea bargains anticipate possible outcomes of jury trials in the event that they fail to reach an agreement. It might even be said that the primary role of a jury trial is allocating bargaining power to each side during a plea bargaining. A trial is less of an end in itself than a means

*Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125. Email: sangmok-at-hss.caltech.edu. I am grateful to Leeat Yariv for encouragement and guidance. Luke Boosey, Kim Border, John Duggan, Federico Echenique, Matias J. Iaryczower, Jean-Laurent Rosenthal, Matthew Shum, and Colin Stewart have made many helpful observations. Morgan Kousser, Hannah Wei, and Caltech undergraduates at the Hixon Writing Center helped me significantly with writing and editing.

[1] In this paper, prosecutors and defendants are all referred to as male, and jurors are all referred to as female.

[2] See table 4.2 in U.S. Department of Justice, Bureau of Justice Statistics, Compendium of Federal Justice Statistics, 2004.

to force negotiations.[3] On the other hand, jurors' behavior in any jury trial that does take place hinges on the outcome of the plea bargaining. Since the incentive to plead guilty increases if the defendant is truly guilty, cases with innocent defendants tend to go to trial. The jury trial incorporates this selection bias in its verdict. Therefore, the jury trial outcomes, which become a basis of pleading decisions, are indeed affected by pleading decisions.

This paper provides a model of the criminal court process focusing on the interaction between plea bargaining and jury trials. We model plea bargaining such that a prosecutor makes a take-it-or-leave-it plea bargain offer and a defendant, who is either guilty or innocent with equal probabilities, pleads either guilty or not guilty. If the defendant pleads guilty, then the case terminates with the offered charge. Otherwise, a jury trial, in which we acquire a framework based on the strategic voting literature, follows. During the process, the defendant attempt to avoid punishments, jurors try to convict the guilty and acquit the innocent, and the prosecutor tries to deliver punishments to the guilty while minimizing the mistake of punishing the innocent.[4] The effect of plea bargaining on jury trials is obvious. When a jury trial takes place, jurors are aware that the defendants whom they face have denied the crime by pleading not guilty. The strategic voting model captures this awareness in terms of jurors' prior belief about how likely a defendant is guilty, which in turn affects jury trial outcomes. The effect of jury trials on plea bargaining is also evident. When the prosecutor and the defendant participate in plea bargaining, they anticipate possible outcomes of the jury trial. In their pleading decisions, the defendant compares the plea bargain offer with possible jury trial outcomes, which the prosecutor already took into account when making the plea bargain offer.

By influencing jurors' belief in the proportion of the guilty defendants, plea bargaining manipulates jurors' voting behavior to resembles the voting behavior in the strategic voting *without* plea bargaining. Yet, the jurors may echo the prosecutor's preferences over mistakenly delivered (or undelivered) punishments to the innocent (or the guilty). The prosecutor's objective is to deliver punishments to guilty defendants, while minimizing the mistake of punishing innocent defendants. To achieve this objective, the prosecutor directly controls the punishment level of the guilty pleas. However, the optimal level is determined by how it will manipulate jurors' behavior, because ex-ante punishment levels are eventually determined by the conviction probabilities in jury trials. In order

[3] Mnookin and Kornhauser (1979) call this "Bargaining in the shadow of the law".

[4] We assume that prosecutors may not single-mindedly pursue convictions, ignoring possible convictions of the innocent. We will justify this assumption in Section 2.

to see this, consider that (1) if the bargain offer is acceptable for the 'guilty,' compared to the jury trial outcome, 'guilty' defendants will plead guilty. Jurors subsequently update their beliefs, reflecting a lower proportion of guilty defendants in jury trial, and consequently lowering the conviction probabilities. As a result, the bargain offer will become unacceptable for 'guilty' defendants. (2) If the bargain offer is unacceptable, the opposite story follows. As the jurors believe that a higher proportion of defendants who come to trial are guilty, the jurors tend to increase their probability of voting for conviction. When this occurs, the bargain offer may become acceptable for the guilty. (3) In general, receiving a guilty plea punishment and undergoing a jury trial will become indifferent for the 'guilty'. (4) Meanwhile, the innocent will not plead guilty because they are less likely be convicted in trial than the guilty, and the guilty are indifferent between pleading guilty and undergoing a jury trial. Thus, the ex-ante punishment for the innocent is also determined by the conviction probability in a jury trial.

Therefore, the prosecutor wants to manipulate jurors' behavior to render the ideal levels of conviction probabilities. The prosecutor cannot force a certain voting behavior to jurors, and their voting behavior will follow an equilibrium behavior. Under this restriction, the ideal jurors' voting behavior will be induced when the jurors' preferences coincide with the prosecutor's preferences. From this observation, the prosecutor controls the proportion of guilty defendants going to jury trials, such that jurors' preferences combined with the distorted proportion coincide with the prosecutors' preference. For instance, as the prosecutor increasingly lowers the guilty plea charge, a higher proportion of guilty defendants plead guilty, and a defendant in a jury trial is more likely to be innocent. Then jurors are more careful when voting to avoid mistakes of convicting the innocent. This manipulated jurors' behavior fits into the prosecutor's preference, when the prosecutor cares more than jurors about mistakenly delivering punishments to innocent defendants. However, such manipulation is possible only in one direction: manipulating jurors to vote more for acquittal. Because the guilty are more likely to take the bargain offer, plea bargaining can only decrease the proportion of the guilty defendants, When the prosecutor cares less about convicting the innocent, and is more averse about acquitting the guilty, plea bargaining is useless in manipulating the jurors' behavior.

Having a combined model of plea bargaining and a jury trial, we can re-interpret implications of the strategic voting literature. As an example, we revisit the comparison of two voting rules, unanimity and non-unanimity, which is studied in Feddersen and Pesendorfer (1998). Feddersen and Pesendorfer find that the unanimity is inferior in terms of the probabilities of convicting the innocent and acquitting the guilty. Under the unanimity, the probabilities do not vanish as the number of jurors grows; whereas, the probabilities vanish under any non-unanimous rule. Here, the convergence results only rely on the voting rules, and the jurors' preferences determine the convergence speeds. One observation in the previous paragraph is that jurors' voting behavior resembles the voting behavior in the strategic voting model without plea bargaining, though preferences may differ. Therefore, with respect to ex-ante delivering punishment of the entire judicial process, inferiority of the unanimity persists in the addition of plea bargaining to the model.

Lastly, this paper sheds light on an economic justification of plea bargaining, signaling, which is *not* motivated by saving trial costs.[5] Indeed, we assume that a trial costs nothing; not only are explicit costs such as time and efforts excluded, but all players are also assumed to be risk neutral. (They are unafraid of uncertainty in trial outcomes.) In this sense, the model is very much like a signaling game. Given the punishment for a guilty plea, a defendant, as a sender, signals his true type by pleading either guilty or not guilty. Afterwords the jurors, as receivers, update their beliefs on the sender's type and determine conviction probabilities. In the prosecutor's viewpoint, plea bargaining allows the court to screen out some guilty defendants before going to a jury trial. Since the accused know whether they are guilty, plea bargaining serves as a self-selection mechanism. As such, it may contribute to the accuracy of the jury trial, on which the entire court performance hinges.

1.2 Related Literature

The related literature can be divided into those two groups: the first group analyzes strategic voting behavior in a jury trial; the second studies either plea bargaining itself or interactions between plea bargaining and jury trials, but without a formal model of jury trials.

A study on collective decision-making under uncertainty is motivated by the Condorcet jury theorem (Condorcet (1785)). Assuming two possible true states, Condorcet models a situation in which a group of people, each of whom is imperfectly and privately informed about the true state, makes a decision by voting for one alternative. Although the members have a common interest in choosing the true state, imperfect private infor-

[5] We do not consider the legal justification of plea bargaining.

mation generates conflict of interests at the time of voting. The theorem says that the group can more efficiently aggregate private information with simple majority rule than if each member acts alone.

The Condorcet jury theorem assumes that each juror votes by following her private information, but recent research illustrates that such action is not consistent with Bayesian Nash equilibrium behavior (Austen-Smith and Banks (1996); Feddersen and Pesendorfer (1996)). The basic intuition departs from the fact that a vote affects a group decision only when the juror is pivotal. A strategic juror incorporates this fact in her voting decision and votes by assuming that she is pivotal. Even when private information is more likely from a certain true state, there are cases where her pivotal state convinces her to follow other jurors' votes against her private information. This strategic voting behavior is evidenced by experimental studies (Guarnaschelli, McKelvey, and Palfrey (2000); Goeree and Yariv (2010)).

Motivated by the strategic voting hypothesis, Feddersen and Pesendorfer (1998) apply the model to a criminal court trial. One of the main findings is the inferiority of the unanimity rule; as the number of jurors grows, the probabilities of convicting the innocent and acquitting the guilty do not vanish under unanimity; whereas, those probabilities converge to zero under all non-unanimous rules. Coughlan (2000) extends the case to mistrial or limited communication among the jurors, and illustrates somewhat divergent results. He points out that a disagreement under unanimity does not automatically yield an acquittal, but rather a mistrial. If a mistrial always results in a new trial, the probability of trial errors is minimized under the unanimity rule. Moreover, assuming that jurors can reveal private signals before the final decision, if sincere revelation and sincere voting behavior is an equilibrium behavior under a non-unanimous rule, they are also in equilibrium under unanimity. Therefore, the unanimity rule may not be inferior. Our research departs from these papers by including plea bargaining. While the previous literature contemplates the consequences of various voting rules in the context of a jury trial, we study the outcomes in the entire judicial process, including both plea bargaining and a jury trial. We find that the addition of plea bargaining to the model preserves the inferiority of the unanimous jury trial rule.

Separately from plea bargaining, jury deliberation is another process known to preserve the inferiority of unanimity voting. Interpreting jury deliberation as a Bayesian communication game, Austen-Smith and Feddersen (2003) and Gerardi and Yariv (2007) find that with jury deliberation, the inferiority of unanimity persists and non-unanimous voting rules generate the same set of equilibrium outcomes. That is, as long as a voting rule is non-unanimous, an exact voting rule is not crucial in the final decision. An experimental study by Goeree and Yariv (2010) confirms that jury deliberation significantly diminishes the differences of various voting mechanisms in their equilibrium outcomes.

Our study generalizes the strategic voting model beyond the jury trial to the criminal court process. In the strategic voting literature, it is conventional to assume that litigation is exogenously given. However, when defendants and prosecutors actively participate in pre-trial stages, the implications of the strategic voting model may not be directly applicable to the entire court process. By attaching a model on plea bargaining to the strategic voting model, we show that the model can be neatly extended to cover the complete judicial process.

Most literature on plea bargaining approaches the process via a 'bargaining' model. A jury trial contains explicit costs, time, and effort; if participants in a plea bargain do not want to bear additional risks, uncertainty in trial outcome is an additional cost. Given such costs, participants in the plea bargain can share a surplus if they reach an agreement. This surplus division is a 'bargaining' problem. A typical model allows either a prosecutor, a defendant, or both to make bargaining offers. Prosecutors know the seriousness of the crime, while the defendant knows whether he is guilty. For a brief summary on this topic, see Cooter and Rubinfeld (1989).

It is undeniable that plea bargaining initially originated as a way of avoiding jury trial costs. However, what we focus on in this paper is the process's welfare effects due to factors other than trial costs, a subject that has received less attention. Grossman and Katz (1983) show that the plea bargain serves as an insurance and screening device. In the former role, it protects the innocent and society against cases where a trial process produces incorrect findings and delivers severe punishments. Although innocent defendants may falsely plead guilty due to the threat of conviction, the sentence will be lenient in such cases. In the latter role, plea bargains sort the guilty and innocent like a self-selection mechanism. Since the mechanism ensures that violators of the law are indeed punished, it may contribute to the accuracy of the legal system. The first role is irrelevant to our model, since we assume that prosecutors and defendants are risk neutral, and consequently need no insurance. The second role shares the same motivation as ours. A major difference from our paper is that Grossman and Katz (1983) lack interactions between plea bargaining and a jury trial. They assume that plea bargaining is a screen device affecting, but never being affected by, the jury trial.

While previous literature studied either plea bargaining assuming an exogenously given trial behavior, or a jury trial assuming an exogenous litigation process, our model allows the plea bargaining and jury trial processes to influence each other in a unified model. In terms of such interaction, Priest and Klein (1984) is one of the studies closest to our paper, because they clarify the relationship between litigation behavior and jurors' behavior in the jury trial. The set of disputes settled and the set litigated are not necessarily the same. Their important assumption is that the potential litigants produce rational estimates of the likely decision by affecting the belief of the jurors. As in our paper, Priest and Klein consider interactions between the pre-trial process and the jury trial. However, while Priest and Klein informally model how biased jurors' beliefs affect the jury decision, we explicitly capture the dynamic by employing a strategic voting model.

2 The Model

A criminal court process begins with a prosecutor indicting a suspect. We assume that the defendant is either guilty (G) or innocent (I), which occur with equal probabilities.

1. Plea Bargaining:

   The prosecutor makes a take-it-or-leave-it plea bargain offer with $\theta \in [0, 1]$ proportion of the original charge. The defendant can plead either *guilty* or *not guilty*. If the defendant pleads guilty, the case terminates and the θ proportion of the original punishment is delivered. Otherwise, the plea bargain is withdrawn, and the case goes to a jury trial. A plea bargain gives the defendant an opportunity to avoid the judgment of conviction.

2. A Jury Trial:

   Our jury model is based on a strategic voting hypothesis in Feddersen and Pesendorfer (1998). A jury consists of $n(n > 1)$ jurors and a voting rule $\hat{k}(0 < \hat{k} \leq n)$. During the trial, each juror interprets testimony by the witnesses. We follow much of the strategic voting literature and describe this interpretation by stating that each juror receives a private signal g or i, which is positively correlated with the true states, as given by

$$Pr[g|G] = Pr[i|I] = p, \quad Pr[i|G] = Pr[g|I] = 1 - p \tag{1}$$

where $p \in (.5, 1)$; a juror has a probability p of receiving a correct signal, and a probability $1 - p$ of receiving an incorrect signal.

The jury reaches a decision by casting votes simultaneously. Each juror can vote for either conviction or acquittal. If the number of conviction votes is larger than the voting rule $\hat{k}$, the defendant is convicted (C). Otherwise, the defendant is acquitted (A). The punishment accompanied by C and A are normalized by 1 and 0 respectively. (Consequently, the punishment by pleading guilty becomes θ.)

Our model assumes that all players behave rationally, where each acts to maximize an appropriately defined utility function. The defendant's utility changes negatively by the amount of punishment; -1 if he is convicted, 0 if he is acquitted, and $-\theta$ if he pleads guilty. He is assumed to be risk neutral; if he perceives that he will be convicted with probability s, then the ex ante utility of going to trial is $s \cdot 1 + (1 - s) \cdot 0$. The defendant wants to minimize punishment and thus maximize his expected utility.

All jurors have identical preferences. We normalize the preferences so that correct judicial decisions incur no utility gains or losses: $u[C|G] = u[A|I] = 0$. Given this normalization, convicting innocent or acquitting guilty defendants incur utility losses, $u[C|I] = -q$ and $u[A|G] = -(1 - q)$, respectively where $q \in [.5, 1)$. We term q as "a level of reasonable doubt." [6] [7]

Finally, we assume that the prosecutor has a preference defined on $[0, 1] \times \{\text{G, I}\}$. Much like the jurors' utilities, when punishment $h \in [0, 1]$ is delivered to a defendant, the prosecutor's utility is given by

$$v[h|I] = -q' \, h \quad , \quad v[h|G] = -(1 - q')(1 - h)$$

where $q' \in [0, 1]$. Prosecutors lose utility if the innocent are punished or the guilty avoid their just punishment. [8]

[6] Suppose that a juror believes that the defendant is guilty with probability $\tilde{q}$. The expected utility from a guilty verdict, $-q(1 - \tilde{q})$, is greater than or equal to the expected utility of an innocent verdict, $-(1 - q)\tilde{q}$, if and only if $\tilde{q} \geq q$. Therefore, when jurors vote for conviction, they use q as the threshold degree of belief that the defendant is guilty. In this respect, Feddersen and Pesendorfer (1998) term q "the threshold level of reasonable doubt."

[7] If $q < 0.5$, we need additional technical conditions to ensure jurors are more likely to vote for conviction when they receive signal g. Even in such case, the analysis in this paper is qualitatively intact.

[8] We may alternatively assume a self-interested prosecutor, who seeks to deliver as much punishment as possible or the highest probability of conviction. However in practice, mistakenly managed cases may later become public, and such exposure will affect a prosecutor's future career. Thus, even a self-

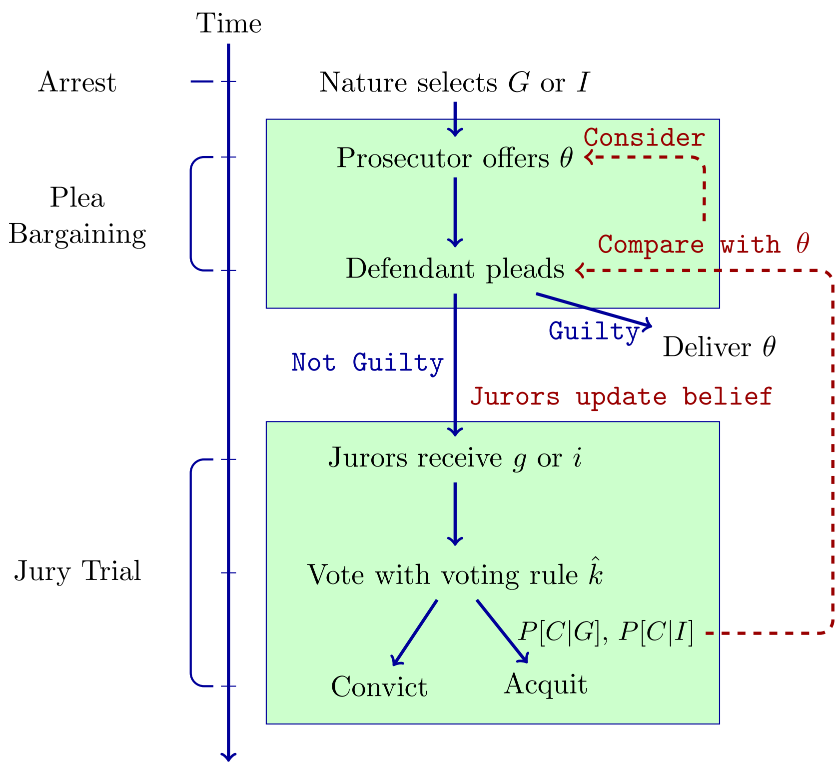


Figure 1: A Criminal Court Process.

Figure 1 summarizes the timing of the model. (1) A prosecutor offers θ as a lesser sentence in a plea bargain. (2) The defendant pleads either guilty or not guilty. (3) If the defendant pleads guilty, a judge respects the bargain, and pronounces sentence θ, and the case terminates. If the defendant pleads not guilty, the case goes to a jury trial. (4) The jury determines whether to convict or acquit. Blue and solid lines in Figure 1 capture how actions at early stages affect actions at later stages; red and dashed lines represent how anticipated outcomes of later stages affect actions at early stages.

3 Jury Trial

Let π denote the jurors' belief in the probability that a defendant is guilty conditioned that the case goes to trial. We assume that a guilty defendant is less likely to go to trial than an innocent are ($\pi \leq .5$). This assumption is innocuous, because the guilty defendant is more likely to send a guilty signal g, and each juror is more likely to vote

interested prosecutor will be concerned with false prosecutions. We represent this concern with flawed cases with a parameterized weight, q'.

for a conviction when she receives a signal g.[9] Thus the guilty have a higher chance of being convicted. As defendants anticipate such jury behavior, the guilty tend to plead guilty, and are therefore less likely to go to trial, relative to the innocent.

A pair (σ_g^j, σ_i^j) in $[0,1] \times [0,1]$ represents a strategy of juror j. Juror j votes for conviction with probability σ_g^j when she receives a signal g, and she votes for conviction with probability σ_i^j if the signal is i. We consider *symmetric equilibrium voting behavior* in which all jurors adopt the same strategy, and denote a symmetric strategy profile as (σ_g, σ_i), without specifying a particular juror. Since the jury trial is modeled as a symmetric game, there exists at least one symmetric equilibrium voting behavior (see Appendix A).[10]

We then find a symmetric voting behavior which gives all jurors the highest expected coordinated payoff. Since all jurors have the same preference for judicial decisions, particularly when convicting the innocent and acquitting the guilty, this is a natural way of refining the symmetric voting behavior. We call this refined behavior an *efficient symmetric voting behavior*, or more succinctly an *efficient voting behavior*.

A rational juror understands that her vote affects the verdict only when she is pivotal.[11] She takes into account not only the private signal (g or i), but also additional information from the event of being pivotal (piv), as evidence of guilt. The juror also knows that defendants in a trial could have pleaded guilty. Thus, the jurors' belief in the probability that a defendant on a trial is guilty (π) also affects her voting behavior.

Let $P[G|piv, g, \pi]$ denote the posterior probability that the defendant is guilty, conditional on receiving signal g and being pivotal:

$$Pr[G|piv, g, \pi] = \frac{\pi \cdot p \cdot Pr[piv|G]}{\pi \cdot p \cdot Pr[piv|G] + (1-\pi) \cdot (1-p) \cdot Pr[piv|I]}$$

Convicting the defendant changes her expected utility by $-q \cdot Pr[I|piv, g, \pi]$, and acquitting changes her utility by $-(1-q) \cdot Pr[G|piv, g, \pi]$. Given all the information available, $Pr[G|piv, g, \pi] > q$ indicates that evidence of guilt is clear enough to exceed

[9] We formally prove this in subsequent paragraphs.

[10] The existence of symmetric equilibrium voting behavior follows very much like the result that a symmetric finite normal form game has a symmetric Nash equilibrium. We leave the formal proof to Appendix A.

[11] Whether a juror is pivotal or not, of course, depends not only on how the other jurors vote but also on the voting rule - unanimity, simple majority, three-fourths, etc.

the level of reasonable doubt (q). The optimal outcome from the juror's viewpoint is to convict; a rational juror will therefore vote for conviction. Whereas, $Pr[G|piv, g, \pi] < q$ indicates that the optimal outcome for the juror is to acquit; a rational juror will vote for acquittal.

Rational jurors will vote by comparing

$$\frac{Pr[\, G \,|\, piv, g, \pi \,]}{Pr[\, I \,|\, piv, g, \pi \,]} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{if the signal is } g,$$

By expanding the above expression, we obtain the following voting criterion for a juror receiving signal g.

$$\frac{Pr[\, piv \,|G]}{Pr[\, piv \,|I]} \frac{p}{1-p} \frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{if the signal is } g. \tag{2}$$

A similar argument applied to a juror receiving signal i, and we obtain

$$\frac{Pr[\, piv \,|G]}{Pr[\, piv \,|I]} \frac{1-p}{p} \frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{if the signal is } i. \tag{3}$$

The left hand side (LHS) is the likelihood ratio of guilty to innocent, given that a juror is pivotal, multiplied by the likelihood ratio of private information (g or i), times the ratio of belief about the defendant's type; the right hand side (RHS) is the ratio of reasonable doubt. If the LHS is larger than the RHS in equation (2), a juror receiving a private signal g has an incentive to vote for conviction; similarly, if the LHS is larger than the RHS in equation (3), a juror receiving a private signal i has an incentive to vote for conviction.

To state the probabilities of being pivotal precisely, let r_G denote the probability of voting for conviction when the defendant is guilty, and r_I be the same probability when the defendant is, instead, innocent. Since a guilty defendant and an innocent defendant send the signal g with probability p and $1-p$ respectively, we obtain

$$r_G = p\sigma_g + (1-p)\sigma_i, \quad r_I = (1-p)\sigma_g + p\sigma_i. \tag{4}$$

When a voting rule requires $\hat{k}$ ($1 \leq \hat{k} \leq n$) number of conviction votes for a guilty verdict, a juror becomes pivotal when $\hat{k}-1$ other jurors vote for conviction. Assuming that $0 < r_I < 1$, voting criterion (2) becomes[12]

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{p}{1-p} \frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \text{ if the signal is } g, \tag{5}$$

and criterion (3) becomes

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{1-p}{p} \frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \text{ if the signal is } i. \tag{6}$$

The above expressions give key intuitions behind the equilibrium restrictions of jurors' behavior in the jury trial. First, the probability of voting for conviction is non-decreasing in the degree of belief π. This is straightforward from the fact that the LHS of both criteria are increasing in π. Second, jurors are more likely to vote for conviction if the defendant is guilty ($r_G > r_I$). Since the LHS of the former criterion is strictly larger than the LHS of the latter criterion, a juror receiving signal g has a greater probability of voting for conviction than a juror receiving a signal i ($\sigma_g > \sigma_i$). Also, a guilty defendant has a higher chance of sending signal g than an innocent defendant, who is more likely to send signal i. Therefore, when a trial has a guilty defendant, jurors tend to vote for conviction ($r_G > r_I$). Third, as the voting rule requires more conviction votes, jurors are more likely to cast conviction votes: the strategy σ_g and σ_i are non-decreasing in $\hat{k}$. Suppose that a strategy profile (σ_g, σ_i) is an equilibrium behavior at π with $\hat{k}$. Considering that the LHS of both voting criteria are increasing in $\hat{k}$, the pair (σ_g, σ_i) can be equilibrium voting behavior with a belief π' ($\pi' < \pi$) when the voting rule requires conviction votes $\hat{k}'$ ($\hat{k}' > \hat{k}$). Equilibrium voting behavior at π with $\hat{k}'$ is then higher than (σ_g, σ_i), because voting behavior is non-decreasing in π.

The equilibrium voting behavior is derived from the voting criteria. Suppose jurors vote for conviction with probabilities of r_I and r_G, where $0 < r_I < r_G < 1$. That is, jurors do not always vote for acquittal ($r_G > r_I > 0$) and do not always vote for conviction ($r_I < r_G < 1$). Since $\sigma_g > \sigma_i$, three cases of strategies are consistent with such jury behavior: $(0 < \sigma_g < 1, \sigma_i = 0)$, $(\sigma_g = 1, 0 < \sigma_i < 1)$, and $(\sigma_g = 1, \sigma_i = 0)$.

[12]When $r_I = 0$ or $r_I = 1$, (5) and (6) are not defined. In Appendix B, we treat these cases separately when we find equilibrium voting behavior.

When jurors receiving signal g use a mixed strategy (the first case), they necessarily have equal preferences on conviction and acquittal. In such instance, the voting criterion with signal g holds with equality, from which we obtain σ_g and consistent levels of π. When a juror receiving signal i uses a mixed strategy (the second case), we obtain σ_i from the equality of voting criterion with signal i. If jurors receiving a signal g vote for conviction and with signal i vote for acquittal (the third case), the juror receiving guilty signal has enough evidence to vote for conviction; whereas, a juror receiving an innocent signal lacks evidence, and thus votes for acquittal. The corresponding inequalities of voting criteria allows us to find the range of π consistent with such strategy profile.

We state the jury behavior in Proposition 1, and relegate details of computing equilibrium voting behavior to Appendix B. It is convenient to introduce a function $\bar{\pi}$ defined as

$$\bar{\pi}(l\ ; p, q) := \frac{1}{\frac{1-q}{q}\left(\frac{p}{1-p}\right)^{l} + 1}, \quad \forall l \in \mathbb{N}$$

which we can rearrange and obtain

$$\left(\frac{p}{1-p}\right)^{l} \frac{\bar{\pi}(l)}{1-\bar{\pi}(l)} = \frac{q}{1-q}. \tag{7}$$

The above equation gives an intuition of which the function $\bar{\pi}$ is defined. The equation is closely related to the voting criteria: $\bar{\pi}$ maps a number of guilty signals (l) to the degree of belief (π), which gives the minimum amount of evidence for a conviction vote. In other words, if a jury has only a single juror who receives multiple signals, $\bar{\pi}(l)$ is the threshold level of the juror's belief such that once the juror gathers l number of guilty signals, the juror votes for conviction.

With the function $\bar{\pi}$, we can intuitively find the range of belief π consistent with different strategies, $(0 < \sigma_g < 1, \sigma_i = 0)$, $(\sigma_g = 1, 0 < \sigma_i < 1)$, and $(\sigma_g = 1, \sigma_i = 0)$. For instance, given a voting rule requiring $\hat{k}$ number of conviction votes, $(\sigma_g = 1, \sigma_i = 0)$ is not an equilibrium behavior for $\pi < \bar{\pi}(2\hat{k} - n)$. To see this, suppose that a juror receives signal g and she turns out to be pivotal; $\hat{k} - 1$ other jurors vote for conviction and $n - \hat{k}$ jurors vote for acquittal. Considering that jurors act $(\sigma_g = 1, \sigma_i = 0)$, $\hat{k} - 1$ conviction votes indicate the same number of guilty signals, and $n - \hat{k}$ acquittal votes indicate the same number of innocent signals. Thus, being pivotal is equivalent to observing $2\hat{k}-n+1$

guilty signals, which results in $2\hat{k}-n$ guilty signals combined with the juror's own guilty signal. When $\pi < \bar{\pi}(2\hat{k}-n)$, $2\hat{k}-n$ guilty signals provides insufficient evidence of guilt. Thus $(\sigma_g = 1, \sigma_i = 0)$ must not be an equilibrium behavior.

Proposition 1 summarizes the equilibrium voting behavior in a jury trial. We say a voting behavior is *reponsive* if the conviction probability with signal g is strictly higher than the probability with signal i. We show that responsive voting is an equilibrium behavior if the beliefs are above a certain threshold level. Intuitively, if responsive voting behavior exists, it must be more efficient than non-responsive behavior, because jurors *use* the private signals in their voting decisions. We confirm that if a responsive voting equilibrium exists, it must be the most efficient voting behavior. Moreover, the conviction probability for the guilty is higher than for the innocent, and the conviction probabilities are non-decreasing in π. This is because guilty and innocent signals are positively correlated with defendants' true types; thus, the conviction probabilities inherit the properties of conviction voting probabilities of each signal g or i.

Proposition 1 ***jurors' behavior in the jury trial***

1. *If $\pi \geq \bar{\pi}(\hat{k})$, then the efficient voting behavior is responsive. Otherwise, the only symmetric voting behavior (which is also efficient) is the one in which no juror votes for conviction.*

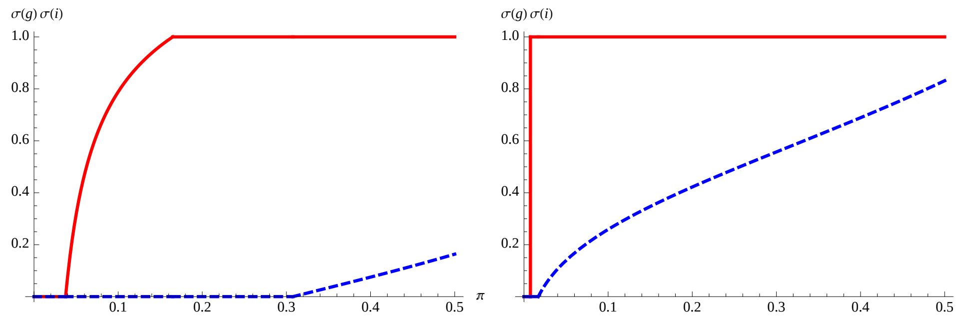
2. *An efficient equilibrium voting behavior (σ_g, σ_i) is non-decreasing in π and $\hat{k}$, and $\sigma_g \geq \sigma_i$.*

3. *Let $\{(P_G, P_I)|\pi\}$ denote the pair of conviction probabilities of the guilty or the innocent, respectively. Convicting the guilty is more likely than convicting the innocent: $P_G \geq P_I$ for all π.*

4. *Let $f_G(\pi) = \{P'_G|\, \exists P'_I,\, (P'_G, P'_I) \in \{(P_G, P_I)|\, \pi\}\}$ and $f_I(\pi) = \{P'_I|\, \exists P'_G,\, (P'_G, P'_I) \in \{(P_G, P_I)|\, \pi\}\}$: correspondences from π to the conviction probabilities of the guilty and the innocent, respectively. Both correspondences are non-decreasing in π: $f_G(\pi) \geq f_G(\pi')$ and $f_I(\pi) \geq f_I(\pi')$ for all $\pi > \pi'$.* [13] [14]

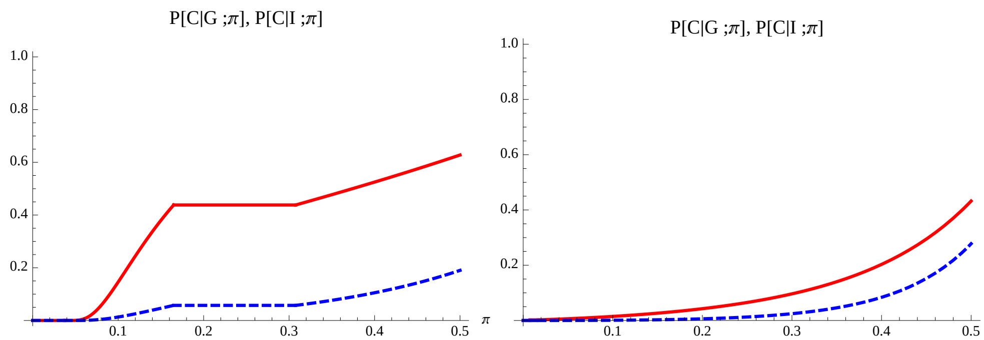
[13] If a real number a is an upper bound of $B \subset \mathbb{R}$, we denote $a \geq B$. If every $a \in A$ is an upper bound of $B \subset \mathbb{R}$, we denote $A \geq B$.

[14] The conviction probabilities, (P_G, P_I), are determined by a voting rule $\hat{k}$, as well as a voting behavior (σ_g, σ_i). Therefore, even though σ_g and σ_i are non-decreasing in $\hat{k}$, the conviction probabilities may not be increasing in $\hat{k}$, because the threshold number of conviction votes also increases.



(a) A Non-unanimous Rule ($\hat{k} = 8$).   (b) The Unanimity Rule ($\hat{k} = 12$).

Figure 2: Efficient symmetric voting behavior with $n = 12$, $p = \frac{6}{10}$, and $q = \frac{1}{2}$



(a) A Non-unanimous Rule ($\hat{k} = 8$).   (b) The Unanimity Rule ($\hat{k} = 12$).

Figure 3: Conviction probabilities with $n = 12$, $p = \frac{6}{10}$, and $q = \frac{1}{2}$

Figure 2 depicts the efficient voting behavior under a non-unanimous rule ($0 < \hat{k} < n$) and the unanimity rule ($\hat{k} = n$) with some parameter values. Red and solid lines represent the probability of voting for conviction with signal g; blue and dashed lines represent the probability of voting for conviction with signal i. Mostly, we have a unique efficient equilibrium voting behavior, except when $\pi = \pi(\hat{k}) = \pi(n)$ under the unanimity rule. The corresponding conviction probabilities are described in Figure 3. Red and solid lines show the conviction probabilities if the defendant is truly guilty; blue and dashed lines show the conviction probabilities of the innocent. Again, we certify that conviction probabilities inherit the properties of conviction voting probabilities: the guilty has a higher chance of being convicted and the conviction probabilities are non-decreasing in π.

4 Plea Bargaining

A prosecutor offers the defendant an opportunity to plead guilty and suffer the penalty $\theta \in [0, 1]$. The defendant has two choices: pleading guilty or not guilty. A guilty plea results in a punishment θ; a not guilty plea sends the case to a jury trial. A guilty defendant compares θ with the conviction probability of the guilty, which we denote as P_G; an innocent defendant compares θ with the conviction probability of the innocent, which we denote as P_I. If θ is larger than P_G, no guilty defendant pleads guilty; similarly, no innocent defendant pleads guilty when θ is larger than P_I.

Such pleading decisions presume that defendants *know* the conviction probabilities of the guilty or the innocent. We can justify this assumption by stating that they rely on defense attorneys. In practice, defendants get advice from defense attorneys, who are aware of whether their previous clients were truly guilty and who can recall the corresponding judicial decisions.[15]

Let ϕ_G and $1 - \phi_G$ denote the probability that a guilty defendant pleads guilty or not guilty, respectively; ϕ_I and $1 - \phi_I$ are defined similarly for an innocent defendant. Recall that π denotes the jurors' beliefs that the defendant is guilty if a case goes to trial. Unless all cases terminate in guilty pleas ($\phi_G > 0$ or $\phi_I > 0$), jurors update their beliefs π by the equation

$$\pi = \frac{1 - \phi_G}{(1 - \phi_G) + (1 - \phi_I)}. \tag{8}$$

If all defendants plead guilty, $\phi_G = \phi_I = 1$, we assume that the jurors update their beliefs by setting them equal to 0.[16]

The relationship between the pleading decisions, ϕ_G and ϕ_I, and the conviction probabilities, P_G and P_I, captures the main interaction between plea bargaining and jury trials. One direction, how pleading decisions affect jury behavior, is explicit. The pleading decisions lead jurors to update their beliefs about the guilt of the defendant (updating π). As we have shown in the previous section, this belief is taken as a part of the evi-

[15] It has been observed that participants in plea bargaining foresee the outcomes of jury trials. In this respect, previous trial decisions significantly influence the parties' bargaining power. Among others, see Bibas (2004) and Stuntz (2004)

[16] Formally, this assumption is equivalent to applying an equilibrium refinement, D1, by Cho and Kreps (1987). Although the details of the refinement are complicated, the intuition is quite simple. Given that the innocent are less likely to be convicted, a not guilty plea guarantees higher utility for the innocent than for the guilty. If a defendant deviates from $\phi_G = \phi_I = 1$ by pleading not guilty, there is a high chance that the defendant is innocent. In such a case, it is reasonable to believe that a deviator must be innocent.

dence of guilt in jury behavior ($\{(P_C, P_I)|\pi\}$). The converse direction, how jury behavior affects pleading decisions, is implicit. The conviction probabilities are taken into account in pleading decisions through the defendants' anticipation (comparing θ vs. P_G or θ vs. P_I). Equilibrium behavior ensures that these interactions must be consistent with each other; the belief π is consistent with pleading decisions ϕ_G and ϕ_I, and the anticipated conviction probabilities are consistent with π ($(P_G, P_I) \in \{(P_C', P_I')|\pi\}$).

Proposition 2 summarizes this equilibrium restriction of the pleading decision and jurors' behavior in the jury trial, given a punishment level θ for pleading guilty. In general, guilty defendants are indifferent between pleading guilty and going to trial ($\theta = P_G$), and the innocent prefer to go to trial ($\theta \geq P_I$).[17] To see why this holds, suppose we have $\theta < P_G$. The guilty will plead guilty, and depending on θ vs. P_I, only the innocent may go to trial. These pleading decisions will lead jurors to believe that all defendants in trials are innocent, and they will vote for acquittal. The corresponding conviction probabilities are zero ($\{(P_G, P_I)|\pi\} = \{(0,0)\}$). Therefore, $\theta < P_G$ must not be an outcome of equilibrium behavior. On the other hand, $\theta > P_G$ can be an equilibrium behavior only if the prosecutor offers relatively high punishment for pleading guilty. In that event, all defendants will go to trial, and if the induced conviction probabilities are still lower than θ, such pleading decisions will be rational.

Proposition 2 ***Pleading Decision and Jury Trial***

Suppose that a prosecutor offers θ as punishment for pleading guilty, and the jury follows an efficient voting behavior. Either one of the following statements, but not both, must be true.

- ($\theta > P_G \geq P_I$)*, which leads all defendants to plead not guilty, and all cases go to jury trial (*$\phi_G = \phi_I = 0$*). The updated belief* π *is equal to the prior probability* $\pi_0 = .5$*, and* $(P_G, P_I) \in \{(P_G', P_I')|.5\}$.
- *The guilty are indifferent between pleading guilty and undergoing a jury trial (*$\theta = P_G$*); innocent defendants weakly prefer to plead not guilty (*$\theta \geq P_G$*). These pleading decisions yield a level of belief* π*, which leads the jury to induce the conviction probabilities of the guilty (*P_G*) such that* $(P_G, P_I) \in \{(P_G', P_I')|\pi\}$ *with some* P_I.

[17] Lemma 5 In Appendix B.3 show that $f_G(\pi)$ is an upper hemicontinuous correspondence with non-empty convex values. If the θ is in $[0, \sup f_G(\pi = .5)]$, then by Intermediate Value Theorem, there exists π such that $\theta = P_G \in f_G(\pi)$.

The prosecutor wants to offer punishment for pleading guilty of θ yielding his highest expected equilibrium payoff. Using the equilibrium restrictions on jury behavior and pleading decisions, the prosecutor's problem is summarized as the following optimization problem.

$$\max_{\theta\in[0,1]} -\frac{1}{2}q'\Big(\phi_I\theta+(1-\phi_I)P_I\Big)-\frac{1}{2}(1-q')\Big(\phi_G(1-\theta)+(1-\phi_G)(1-P_G)\Big) \tag{9}$$

$$\text{such that}\quad \begin{array}{ll} (a.1) & \phi_G \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_G \\ (a.2) & \phi_I \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_I \\ (b) & \pi = \begin{cases} 0 & \text{if } \phi_G=\phi_I=1 \\ \frac{1-\phi_G}{(1-\phi_G)+(1-\phi_I)} & \text{otherwise.} \end{cases} \\ (c) & (P_G,P_I)\in\{(P'_G,P'_I)|\pi\}. \end{array}$$

The objective function is the prosecutor's expected utility function. The prosecutor's utility is decreased by q' if the innocent are mistakenly punished. The punishment is either as a result of a guilty plea, $\phi_I\,\theta$, or of conviction in jury trial, $(1-\phi_I)P_I$. When the guilty go without being fully punished, it also decreases the prosecutor's utility by $(1-q')$. Such case is either as a result of a guilty plea, $\phi_G(1-\theta)$, or of acquittal in a jury trial, $(1-\phi_G)(1-P_G)$.

The prosecutor anticipates that the defendants will make rational pleading decisions and the jurors' behavior will follow equilibrium voting behavior, which restrict the prosecutor's optimization: $(a.1)$ and $(a.2)$ represent pleading decisions by guilty and innocent defendants, respectively; (b) captures the notion that jurors rationally update their belief π to be consistent with pleading decisions; (c) requires that defendants rationally anticipate jury behavior, and consequently its conviction probabilities; and $\{(P_G,P_I)|\pi\}$ in (c) presumes that jurors will follow the efficient voting behavior.

Solving the prosecutor's problem requires mathematical techniques, which we leave to Appendix C, but the motivation behind the prosecutor's optimal level of θ is quite intuitive. To illustrate the main idea, we show that the prosecutor is mainly concerned with how to manipulate the jurors' belief π.

For the restrictions in (9), the prosecutor only needs to focus on the second case in Proposition 2. To see why the second restriction contains the first, suppose that the second restriction holds in an equilibrium. The punishment following a guilty plea is so high that all defendants go to trial. The corresponding prosecutor's utility can also be achieved by offering $\theta = \bar{\theta}$ where $\bar{\theta} := \sup f_G(.5)$. Regardless of whether a guilty person pleads guilty or not guilty, the prosecutor achieves the same utility gain or loss, which leads to the same utility change in the case of all defendants pleading not guilty.

Using the first equilibrium restriction in Proposition 2, we can simplify the prosecutor's objective function in (9). Unless $\theta = 0$, we have $\theta = P_G > 0$. The efficient voting behavior becomes responsive ($P_G > P_I$), and all innocents go to trial ($\phi_I = 1$). Then the prosecutor's objective function becomes

$$-\frac{1}{2}q' P_I - \frac{1}{2}(1-q')(1-P_G). \tag{10}$$

We now see that the prosecutor's main concern is to manipulate the jurors' belief π, thereby leading to the most preferable jury behavior. One thing to note here is that the prosecutor is not allowed to force jurors to take a certain voting strategy. That is, he can at best lead to one of the efficient voting behaviors.

To see *how* the prosecutor should manipulate the jurors' belief π, we revisit the jurors' voting criteria. We have shown that jurors vote for conviction by comparing

$$\frac{Pr[\,piv\,|G]}{Pr[\,piv\,|I]}\,\frac{p}{1-p}\,\frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{if the signal is } g,$$

and

$$\frac{Pr[\,piv\,|G]}{Pr[\,piv\,|I]}\,\frac{1-p}{p}\,\frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{if the signal is } i,$$

We can modify the expressions into

$$\frac{Pr[\,piv\,|G]}{Pr[\,piv\,|I]}\,\frac{p}{1-p}\,\frac{.5}{1-.5} \quad \text{vs} \quad \frac{q}{1-q}\,\frac{1-\pi}{\pi} \quad \text{if the signal is } g,$$

and

$$\frac{Pr[\,piv\,|G]}{Pr[\,piv\,|I]}\,\frac{1-p}{p}\,\frac{.5}{1-.5} \quad \text{vs} \quad \frac{q}{1-q}\,\frac{1-\pi}{\pi} \quad \text{if the signal is } i.$$

The above two versions of the voting criteria lead to the same voting behavior; jurors receiving signal g (or i) vote for conviction if confronted with the former pair of criteria if and only if jurors receiving signal g (or i) vote for conviction if confronted with the latter pair of criteria. That is, the jury behavior with a belief π and the ratio of reasonable doubts $\frac{q}{1-q}$ is equal to the jury behavior with a belief .5 and the ratio of reasonable doubt equal to $\frac{q}{1-q}\frac{1-\pi}{\pi}$. Consequently, we can reinterpret the prosecutor's effort to manipulate the jurors' beliefs as an effort to change the level of the jurors' reasonable doubt, while fixing the belief equal to π_0, or .5. The question, "how to manipulate the jurors' belief", is then the same as, "which level of the jurors' manipulated reasonable doubt is the most preferable to the prosecutor."

Intuitively, the prosecutor prefers to have jurors manipulated reasonable doubt to coincide perfectly with his preference weights on mistakenly delivered or undelivered punishments: $\frac{q'}{1-q'} = \frac{q}{1-q}\frac{1-\pi}{\pi}$. However, the prosecutor can manipulate the jurors' reasonable doubt in only one direction; he can only increase the reasonable doubt by inducing $\pi \leq .5$. When the jurors, rather than the prosecutor, care about punishing the innocent ($q > q'$), the prosecutor has no choice but to induce $\pi = .5$ by offering $\theta \geq \bar{\theta}$. The following proposition presents the prosecutor's optimal behavior under an equilibrium constraint. (The formal proof is in the Appendix D.)

Proposition 3 ***Equilibrium behavior in the Criminal Court.***

1. *If $q < q'$, the prosecutor offers a punishment for pleading guilty such that some guilty defendants and none of the innocent defendants plead guilty. The jury's behavior is the same as the jury behavior of the jury trial model without plea bargaining; however, the jurors behave as if they have the prosecutor's preference, q'.*

2. *If $q \geq q'$, the prosecutor offers a harsh punishment for pleading guilty ($\theta \geq \bar{\theta}$), and all defendants plead not guilty and go to trials. The jury behavior is the same as the jury behavior of the jury model without plea bargaining.*

5 Comparison of Alternative Voting Rules.

As a direct application of Proposition 3, we can re-examine the previous findings from the jury model. Feddersen and Pesendorfer (1998) find that the unanimity rule is inferior to non-unanimous rules. As the number of juror gets large, the chance of convicting the innocent and the chance of acquitting the guilty do not converge to zero under the unanimous rule; whereas, both converge to zero if the voting rule is non-unanimous. Assuming that the jury trial employs either the unanimity or a non-unanimous rule, we reconfirm that the previous results are robust to plea bargaining.

One of the main conclusions in Proposition 3 is that jury behavior in a criminal court with plea bargaining is similar to equilibrium behavior in a jury model without plea bargaining. If $q > q'$, the behaviors are exactly same; if $q \leq q'$, we can mimic the jury behavior with plea bargaining in a jury model without plea bargaining by assuming that jurors echo the prosecutor's preference. Thus, the qualitative findings concerning jury behavior using unanimous and non-unanimous rules are not changed by plea bargaining. Plea bargaining only affects quantitative analyses.

Propositions 2 and 3 in Feddersen and Pesendorfer (1998) are transformed into Corollary 4. The main theme is preserved, but the limiting values are changed. It is worth stressing that while the previous literature considers jury trial outcomes, or conviction probabilities, we treat the outcomes of the entire judicial process: conviction probabilities and punishment for pleading guilty. Therefore, Corollary 4 compares expected punishments, rather than conviction probabilities, under either the unanimity or non-unanimous rules.

Corollary 4 ***The unanimity rule vs. Non-unanimous rules.***

Suppose that a criminal court has a plea bargain and a jury trial with n number of jurors. If the jury requires n votes for conviction, we call the rule unanimity. Otherwise, the jury requires $\hat{k} = \alpha n$ ($0 < \alpha < 1$), and we call the rule non-unanimous.

- *If a jury trial uses the unanimity rule, the expected punishment of the guilty converges to $1 - \left(\frac{(1-\tilde{q})(1-p)}{\tilde{q}p}\right)^{1-\frac{p}{2p-1}}$ as $n \to \infty$, where $\tilde{q} = \max\{q, q'\}$; for the innocent, it converges to $\left(\frac{(1-\tilde{q})(1-p)}{\tilde{q}p}\right)^{\frac{p}{2p-1}}$.*
- *If the jury trial uses a non-unanimous rule with a fixed α, the expected punishment for the guilty converges to one as $n \to \infty$. The expected punishment for the innocent converges to zero.*

6 Conclusion

We study a criminal court process where plea bargaining interacts with jury trial. A plea bargain is initiated when a prosecutor offers a take-it-or-leave-it guilty pleading punishment. Then, a guilty or an innocent defendant will choose to plead either guilty or not guilty. Pleading not guilty is followed by a jury trial in which jurors are assumed to follow strategic voting behavior. The strategic voting model allows us to obtain an explicit dynamic of which pleading decisions affect jurors' beliefs, thereby influencing their voting behavior. This dynamic is also applied how defendants anticipate trial outcomes, with which the defendants compare the guilty plea punishment.

We find that, even if only a proportion of defendants go to jury trial, jury behavior still takes a fundamental role in the court process. If the prosecutor cares about convicting the innocent more than jurors do, then the expected degrees of punishment for the guilty and the innocent in our model are equal to the conviction probabilities of the guilty and the innocent in the jury model without plea bargaining. If the prosecutor cares about convicting the innocent less than jurors do, the expected degrees of punishment for the guilty and the innocent in our model are also equal to the conviction probabilities in the jury model without plea bargaining, but jurors echo the prosecutor's preferences. We also find that the inferiority of the unanimity rule persists in plea bargaining. As the number of jurors grows, the amount of expected false punishment, or delivering punishment to the innocent and not delivering it to the guilty, do not vanish under unanimity. In contrast, they converge to zero under every non-unanimous rule.

This study enlarges the field of strategic voting model beyond the jury trial to the criminal court system. While most previous studies focus on the trial itself, our paper suggests that we can obtain the implications of the strategic voting behavior broadly from the entire court process. In addition, our model provides a framework to study the economic justification of plea bargaining. We also provide formal dynamics implied in the previous literature on the interaction between pre-trial and jury trial. Although we focus on criminal courts, similar motivations are applicable to civil courts with litigation, committees setting an agenda and voting to pass, or even journal referees, when authors try their luck in submitting articles.

In this paper, we manage to simplify the model enough to be analytically tractable. However, this simplification leaves out several interesting issues. An immediate extension is about a prosecutor who is also informed about the defendant's type. If he acquires the information of a defendant, the prosecutor may offer differentiated guilty plea punishments or sometimes discharge cases. The current model can easily be modified to ask these questions. However, due to the binary nature of defendants (either guilty or innocent), the equilibrium analysis yields no interesting implications beyond equilibrium computations. Once we have a richer model accounting for the intensity of degree of guilt, the setup with an informed prosecutor would be an interesting research question.

Appendix A Existence of a symmetric voting equilibrium.

Let $S := \{c, a\} \times \{c, a\}$ be the set of pure strategies: 'c' represents voting for conviction and 'a' for acquittal. A generic strategy $s \in S$ is a pair (s_g, s_i) consisting of voting decisions with signal g and i. Let $\Sigma := \Delta(\{c, a\}) \times \Delta(\{c, a\})$. A generic mixed strategy $\sigma = (\sigma_g, \sigma_i) \in \Sigma$ consists of probabilities of conviction voting with signal g and i. Define continuous functions $u_g(\sigma'_g, \sigma)$ or $u_i(\sigma'_i, \sigma)$ as a juror's expected utility when she receives signal g or i respectively and uses strategy σ', while all other jurors use strategy σ. u_g and u_i are continuous in σ' and σ.

We proceed similar to the existence proof of Nash equilibrium in Nash (1951). For each pure strategy $s \in S$, define a continuous function h as

$$h^s(\sigma) := \big(\max\{\,0\,,\,u_g(s_g, \sigma) - u_g(\sigma_g, \sigma)\}\,,\,\max\{\,0\,,\,u_i(s_i, \sigma) - u_i(\sigma_i, \sigma)\}\big).$$

For each $s \in S$, define a continuous function as

$$y^s(\sigma) := \left(\frac{\sigma_{g:s_g} + h_1^s(\sigma)}{1 + \sum_{t \in \{c,a\}} h_1^t(\sigma)}\,,\,\frac{\sigma_{g:s_i} + h_2^s(\sigma)}{1 + \sum_{t \in \{c,a\}} h_2^t(\sigma)}\right)$$

where $\sigma_{g:s_g}$ and $\sigma_{g:s_i}$ are the probabilities that the mixed strategy $\sigma = (\sigma_g, \sigma_i)$ assigns to each pure strategy s_g and s_i.

The set of functions $y^s(\cdot)$ $\forall s \in S$ defines a mapping from the set of mixed strategy to itself. Similar to the existence proof of Nash equilibrium, a fixed point of $y(\cdot)$ is a symmetric Bayesian Nash Equilibrium (a symmetric equilibrium voting behavior). Since the set of mixed strategy is compact and convex, $y(\cdot)$ has a fixed point by the Brouwer fixed point theorem.

Appendix B Proof of Proposition 1

For each level of belief π, we first find all symmetric equilibrium voting behaviors. Then we compare the jurors' expected payoffs and take the most efficient symmetric voting behavior.

B.1 Finding all symmetric equilibrium voting behaviors.

B.1.1 Non-responsive equilibrium voting behavior

When $\hat{k} < n$, $(\sigma_g = 1, \sigma_i = 1)$ is an equilibrium voting behavior; given that other jurors always vote for conviction, a juror is never pivotal. (Her vote never changes the judicial decisions.) In such case, no juror has an incentive to change her voting strategy from $(\sigma_g = 1, \sigma_i = 1)$. Similarly, $(\sigma_g = 0, \sigma_i = 0)$ is an equilibrium voting behavior when $1 < \hat{k}$.

When $\hat{k} = n$, $(\sigma_g = 1, \sigma_i = 1)$ is not an equilibrium. Given that other jurors always vote for conviction, being pivotal does not give any additional information. Jurors then fully rely on their own private signals. If a juror receives innocent signal, then she compares

$$\frac{1-p}{p}\frac{\pi}{1-\pi} \quad \text{vs.} \quad \frac{q}{1-q}.$$

Note that the evidence innately supports for innocent defendants ($\frac{1-p}{p} < 1$ and $\frac{\pi}{1-\pi} \leq 1$), and reasonable doubt is in favor of acquittal ($\frac{q}{1-q} \geq 1$). A jurors receiving innocent signal does not have enough evidence to vote for conviction; $\sigma_i = 1$ is not a best response to $(\sigma_g = 1, \sigma_i = 1)$. In a similar fashion, when $\hat{k} = 1$, $(\sigma_g = 0, \sigma_i = 0)$ is an equilibrium voting behavior only if $\pi \leq \bar{\pi}(1)$. Being pivotal does not provide any additional evidence, and a juror compares her private signal (g or i), belief (π), and reasonable doubt (q). If the belief π is low, even a guilty signal gives insufficient evidence for conviction voting.

B.1.2 Responsive equilibrium voting behavior

A responsive voting strategy has $0 < \sigma_g$ and $\sigma_i < 1$. We define r_G and r_I as conviction probabilities of guilty and innocent defendants, computed as

$$r_G = p\sigma_g + (1-p)\sigma_i, \quad r_I = (1-p)\sigma_g + p\sigma_i$$

When the jury follows responsive voting behavior, the jury does not always convict nor acquit defendants ($0 < r_G, r_I < 1$). In such case, voting criteria, (5) and (6), are well defined. For convenience, we reproduce the criteria below.

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{p}{1-p} \frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{for signal } g,$$

and

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{1-p}{p} \frac{\pi}{1-\pi} \quad \text{vs} \quad \frac{q}{1-q} \quad \text{for signal } i.$$

We consider each case of strategy and find necessary levels of belief π consistent with the strategy as an equilibrium voting behavior.

Case 1: $0 < \sigma_g < 1$, $\sigma_i = 0$

Conviction and acquittal must be indifferent to a juror receiving signal g. That is

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{p}{1-p} \frac{\pi}{1-\pi} = \frac{q}{1-q}$$

Substituting in $r_G = p\,\sigma_g$ and $r_I = (1-p)\,\sigma_g$, we obtain

$$\left(\frac{1-p\sigma_g}{1-(1-p)\sigma_g}\right)^{n-\hat{k}} \left(\frac{p}{1-p}\right)^{\hat{k}} \frac{\pi}{1-\pi} = \frac{q}{1-q}. \tag{11}$$

Under unanimity ($\hat{k} = n$), the first term in LHS is equal to 1, and the equality holds when $\pi = \bar{\pi}(\hat{k}) = \bar{\pi}(n)$. Then, any $\sigma_g \in (0,1)$ with $\sigma_i = 0$ is an equilibrium voting behavior.

Consider a non-unanimous voting rule $\hat{k}$ ($\hat{k} < n$). Since $\frac{1-p\sigma_g}{1-(1-p)\sigma_g}$ is strictly decreasing in σ_g, by plugging $\sigma_g = 0$ and $\sigma_g = 1$ in (11), we can check that $\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(2\hat{k}-n)$ is necessary for $(0 < \sigma_g < 1, \sigma_i = 0)$ to be an equilibrium voting behavior. Moreover, at most one value of σ_g satisfies the equality. By algebraic manipulation of (11), we find $(\sigma_g, \sigma_i = 0)$ is an equilibrium voting strategy with

$$\sigma_g(\pi) = \frac{\psi_1 - 1}{(1-p)\psi_1 - p} \quad \text{where} \quad \psi_1 = \left(\frac{1-p}{p}\right)^{\frac{\hat{k}}{n-\hat{k}}} \left(\frac{q}{1-q}\,\frac{1-\pi}{\pi}\right)^{\frac{1}{n-\hat{k}}} \tag{12}$$

Case 2: $\sigma_g = 1$, $\sigma_i = 0$

A juror receiving signal g prefers conviction, whereas a juror receiving signal i prefers acquittal. Substituting in $r_G = p$ and $r_I = 1-p$ to voting criteria (5) and (6), we obtain

$$\left(\frac{p}{1-p}\right)^{2(\hat{k}-1)-n} \leq \frac{q}{1-q}\,\frac{1-\pi}{\pi} \leq \left(\frac{p}{1-p}\right)^{2\hat{k}-n} \tag{13}$$

The first inequality is from the criterion with signal i, and the second inequality is from the criterion with signal g. The above inequality is equivalent to $\bar{\pi}(2\hat{k}-n) \leq \pi \leq \bar{\pi}(2(\hat{k}-1)-n)$. When π is between $\bar{\pi}(2\hat{k}-n)$ and $\bar{\pi}(2(\hat{k}-1)-n)$, $(\sigma_g = 1, \sigma_i = 0)$ is an equilibrium voting behavior: every juror follows her own signal.

Case 3: $\sigma_g = 1$, $0 < \sigma_i < 1$

Jurors receiving signal i treat conviction and acquittal equally. That is

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}}\,\frac{1-p}{p}\,\frac{\pi}{1-\pi} = \frac{q}{1-q}$$

Substituting in $r_G = p + (1-p)\sigma_i$ and $r_I = (1-p) + p\sigma_i$, we get

$$\left(\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}\right)^{\hat{k}-1}\left(\frac{1-p}{p}\right)^{n-\hat{k}+1}\frac{1-\pi}{\pi} = \frac{q}{1-q} \tag{14}$$

Note that $\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}$ is strictly decreasing in σ_i. By plugging in $\sigma_i = 0$ and $\sigma_i = 1$, we can verify that $\bar{\pi}(2(\hat{k}-1)-n) < \pi \leq .5$ is necessary if $\sigma_g = 1$ and $0 < \sigma_i < 1$ is an equilibrium voting behavior.

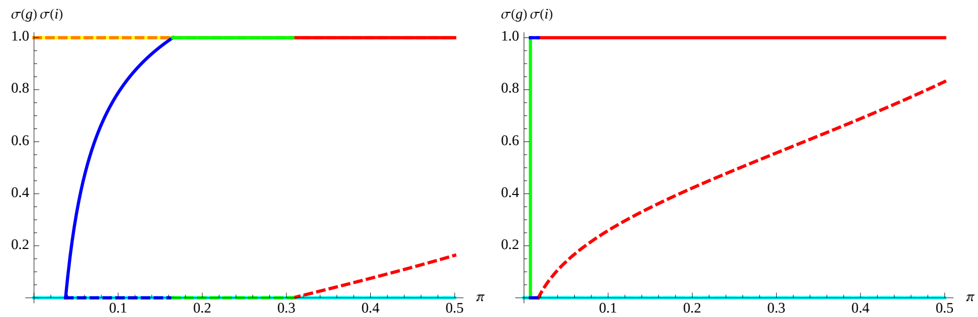
For each level of belief π such that $\bar{\pi}(2(\hat{k}-1)-n) < \pi < .5$, at most one σ_i satisfies the equality. This σ_i combined with $\sigma_g = 1$ forms a symmetric equilibrium voting behavior, and the σ_i is determined as

$$\sigma_i(\pi) = \frac{p - \psi_2(1-p)}{p\,\psi_2 - (1-p)} \quad \text{where} \quad \psi_2 = \left(\frac{p}{1-p}\right)^{\frac{n-\hat{k}+1}{\hat{k}-1}}\left(\frac{q}{1-q}\,\frac{1-\pi}{\pi}\right)^{\frac{1}{\hat{k}-1}} \tag{15}$$

Table 1 summarizes all symmetric equilibrium voting behavior. And Figure 4 illustrates equilibrium voting behaviors with $n = 12$, $p = \frac{6}{10}$, and $q = \frac{6}{10}$, when voting rules are $\hat{k} = 8$ and $\hat{k} = 12$. We used solid lines for σ_g and dashed lines for σ_i. For each π, σ_g and σ_i in a strategy profile (σ_g, σ_i) share the same color. In this example, we observe all three equilibrium cases, but we may not observe some cases under other parameter values. For instance, $\bar{\pi}(2(\hat{k}-1)-n)$, one of the threshold level of belief, may not be

Non-unanimous rules		The unanimity rule	
Non-responsive voting			
$\forall\ \pi$	$(\sigma_g = \sigma_i = 1)$	$\forall\ \pi$	$(\sigma_g = \sigma_i = 0)$
If $\hat{k} > 1$, $\forall\ \pi$	$(\sigma_g = \sigma_i = 0)$		
Responsive voting			
$\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(2\hat{k} - n)$	$(0 < \sigma_g < 1, \sigma_i = 0)$	$\pi = \bar{\pi}(n)$	$(0 < \sigma_g < 1, \sigma_i = 0)$
$\bar{\pi}(2\hat{k} - n) \leq \pi \leq \bar{\pi}(2(\hat{k} - 1) - n)$	$(\sigma_g = 1, \sigma_i = 0)$	$\bar{\pi}(n) \leq \pi \leq \bar{\pi}(n - 2)$	$(\sigma_g = 1, \sigma_i = 0)$
$\bar{\pi}(2(\hat{k} - 1) - n) < \pi \leq .5$	$(\sigma_g = 1, 0 < \sigma_i < 1)$	$\bar{\pi}(2n - 2) < \pi \leq .5$	$(\sigma_g = 1, 0 < \sigma_i < 1)$

Table 1: Symmetric voting equilibrium behavior in jury trial.



(a) A non-unanimous rule ($\hat{k} = 8$).

(b) The unanimity rule ($\hat{k} = 12$).

Figure 4: Symmetric equilibrium voting behavior with $n = 12$, $p = \frac{6}{10}$, and $q = \frac{6}{10}$

defined or may be larger than .5. In such case, $(\sigma_g = 1, \sigma_i = 0)$ is not an equilibrium voting behavior for any $\pi \in [0, .5]$.

B.2 Finding an efficient voting behavior.

For each belief π, there may be several symmetric equilibrium voting behaviors. If a responsive voting exists, intuitively it must be more efficient than non-responsive voting, because jurors essentially *use* private signals to form judgements. We confirm this intuition by comparing responsive voting outcomes with non-responsive voting outcomes. If there is no responsive voting for a π, then one of the non-responsive equilibria, $(\sigma_g = 1, \sigma_i = 1)$ or $(\sigma_g = 0, \sigma_i = 0)$, is efficient.

Given a belief π, conviction probabilities (P_G, P_I) change the jurors' expected payoff by

$$-q \cdot (1 - \pi) \cdot P_I - (1 - q) \cdot \pi \cdot (1 - P_G).$$

The first term corresponds to mistakenly convicting the innocent, and the second term corresponds to mistakenly acquitting the guilty.

Between two non-responsive voting behaviors, ($\sigma_g = \sigma_i = 0$) and ($\sigma_g = \sigma_i = 1$), the former gives a higher jurors' expected utility than the latter, because $q\,(1-\pi)$ is larger than $(1-q)\,\pi$.

When $\pi < \bar{\pi}(\hat{k})$, there is no responsive voting behavior, thus ($\sigma_g = \sigma_i = 0$) is the efficient equilibrium voting behavior. When $\pi \geq \bar{\pi}(\hat{k})$, there is a responsive voting behavior, and the responsive voting is more efficient than ($\sigma_g = \sigma_i = 0$) if and only if the conviction probabilities (P_G, P_I) of responsive voting satisfies

$$-q\ (1-\pi)\ P_I - (1-q)\ \pi\ (1-P_G) > -(1-q)\ \pi$$

which we can rewrite as

$$\frac{P_G}{P_I} = \frac{\sum_{j=\hat{k}}^{n} \binom{n}{j} r_G^j (1-r_G)^{n-j}}{\sum_{j=\hat{k}}^{n} \binom{n}{j} r_I^j (1-r_I)^{n-j}} > \frac{q}{1-q}\frac{1-\pi}{\pi}. \tag{16}$$

If the above inequalities hold as equalities, then responsive voting behavior and ($\sigma_g = 0, \sigma_i = 0$) are both equally efficient.

We proceed separately with non-unanimous rules and the unanimity rule.

B.2.1 Non-unanimous rules ($\hat{k} < n$)

In order to verify (16), first note that $k' > k$, $r_G > r_I > 0$ implies

$$\frac{r_G^{k'}(1-r_G)^{n-k'}}{r_I^{k'}(1-r_I)^{n-k'}} > \frac{r_G^{k}(1-r_G)^{n-k}}{r_I^{k}(1-r_I)^{n-k}}. \tag{17}$$

Also note that

$$\text{if } x, x' > 0 \text{ and } y, y' > 0, \qquad \frac{x'}{y'} > \frac{x}{y} \quad \text{implies} \quad \frac{x+x'}{y+y'} > \frac{x}{y}. \tag{18}$$

Sequentially applying (17) to LHS of (16) using (18), we obtain

$$\frac{\sum_{j=\hat{k}}^{n} \binom{n}{j} r_G^j (1-r_G)^{n-j}}{\sum_{j=\hat{k}}^{n} \binom{n}{j} r_I^j (1-r_I)^{n-j}} > \frac{r_G^{\hat{k}}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}}(1-r_I)^{n-\hat{k}}}.$$

Therefore, to prove (16), it is enough to show

$$\frac{r_G^{\hat{k}}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}}(1-r_I)^{n-\hat{k}}} \geq \frac{q}{1-q}\,\frac{1-\pi}{\pi}. \tag{19}$$

We proceed with each case of responsive voting behavior.

Case 1: $(0 < \sigma_g < 1, \sigma_i = 0)$, where $\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(2\hat{k}-n)$.

By substituting in $r_G = p\sigma_g$ and $r_I = (1-p)\sigma_g$, the LHS of (19) becomes

$$\frac{r_G^{\hat{k}}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}}(1-r_I)^{n-\hat{k}}} = \left(\frac{1-p\sigma_g}{1-(1-p)\sigma_g}\right)^{n-\hat{k}}\left(\frac{p}{1-p}\right)^{\hat{k}}.$$

An equilibrium restriction (11) implies that the RHS of the above expression is equal to the RHS of (19). Thus (19) holds under equality.

Case 2: $(\sigma_g = 1, \sigma_i = 0)$, where $\bar{\pi}(2\hat{k}-n) \leq \pi \leq \bar{\pi}(2(\hat{k}-1)-n)$.

Since $r_G = p$ and $r_I = 1-p$, the LHS of (19) is

$$\frac{r_G^{\hat{k}}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}}(1-r_I)^{n-\hat{k}}} = \left(\frac{p}{1-p}\right)^{2\hat{k}-n}.$$

From (13), equation (19) must be true.

Case 3: $(\sigma_g = 1, 0 < \sigma_i < 1)$, where $\bar{\pi}(2(\hat{k}-1)-n) < \pi \leq .5$.

Note that (14) is a necessary equilibrium restriction. Since $\pi \leq .5$ and $p > .5$,

$$\left(\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}\right)^{\hat{k}-1}\left(\frac{1-p}{p}\right)^{n-\hat{k}+1} = \frac{q}{1-q}\,\frac{1-\pi}{\pi}$$

By substituting in $r_G = p + (1-p)\,\sigma_i$, $r_I = (1-p) + p\,\sigma_i$, we obtain

$$\frac{r_G^{\hat{k}}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}}(1-r_I)^{n-\hat{k}}} = \left(\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}\right)^{\hat{k}}\left(\frac{1-p}{p}\right)^{n-\hat{k}} \geq \left(\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}\right)^{\hat{k}-1}\left(\frac{1-p}{p}\right)^{n-\hat{k}+1}$$

Inequality (19) is derived from the above two inequalities.

B.2.2 The unanimity rule ($\hat{k} = n$)

If the voting rule follows unanimity, then (16) becomes

$$\frac{P_G}{P_I} = \left(\frac{r_G}{r_I}\right)^n > \frac{q}{1-q}\frac{1-\pi}{\pi}. \tag{20}$$

If the above inequality holds, responsive voting is more efficient than $(\sigma_g = 0, \sigma_i = 0)$; if LHS and RHS are equal, both responsive and $(\sigma_g = 0, \sigma_i = 0)$ are equally efficient.

Case 1: $(0 < \sigma_g < 1, \sigma_i = 0)$, where $\pi = \bar{\pi}(n)$.

By substituting in $r_G = p\sigma_g$ and $r_I = (1-p)\sigma_g$, the LHS of (20) becomes

$$\left(\frac{r_G}{r_I}\right)^n = \left(\frac{p}{1-p}\right)^n.$$

By definition of $\bar{\pi}(\cdot)$ and $\pi = \bar{\pi}(n)$, (20) holds as an equality. Thus, both $(0 < \sigma_g < 1, \sigma_i = 0)$ and $(\sigma_g = 0, \sigma_i = 0)$ are equally efficient.

Case 2: $(\sigma_g = 1, \sigma_i = 0)$, where $\bar{\pi}(2\hat{k} - n) \leq \pi \leq \bar{\pi}(2(\hat{k} - 1) - n)$.

Since $r_G = p$ and $r_I = 1 - p$, the LHS of (20) is

$$\left(\frac{r_G}{r_I}\right)^n = \left(\frac{p}{1-p}\right)^n.$$

By definition of $\bar{\pi}(\cdot)$, (20) holds as an equality when $\pi = \bar{\pi}(2\hat{k} - n) = \bar{\pi}(n)$; otherwise if $\bar{\pi}(n) < \pi \leq \bar{\pi}(2(\hat{k} - 1) - n)$ then (20) holds with a strict inequality. Thus, when $\pi = \bar{\pi}(n)$, both $(\sigma_g = 1, \sigma_i = 0)$ and $(\sigma_g = 0, \sigma_i = 0)$ are equally efficient; when $\bar{\pi}(n) < \pi \leq \bar{\pi}(2(\hat{k} - 1) - n)$, responsive voting $(\sigma_g = 1, \sigma_i = 0)$ is the most efficient.

Case 3: $(\sigma_g = 1, 0 < \sigma_i < 1)$, where $\bar{\pi}(2(\hat{k} - 1) - n) < \pi \leq .5$.

By substituting in $r_G = p + (1-p)\,\sigma_i$, $r_I = (1-p) + p\,\sigma_i$, we obtain

$$\left(\frac{r_G}{r_I}\right)^n = \left(\frac{p + (1-p)\sigma_i}{(1-p) + p\sigma_i}\right)^n > \left(\frac{p + (1-p)\sigma_i}{(1-p) + p\sigma_i}\right)^{n-1} \frac{p}{1-p} = \frac{q}{1-q}\frac{1-\pi}{\pi}$$

where the last equality is from the voting criterion (14). Responsive voting is the most efficient.

B.3 Proof of each item in Proposition 1.

We have shown that there exists a responsive equilibrium voting behavior if the belief π is no less than $\bar{\pi}(\hat{k})$. Moreover, responsive voting yields the highest jurors' expected utility (Item 1).[18]

From the closed form solution of responsive voting behavior, we observe that σ_g and σ_i are constant on $[0, \bar{\pi}(\hat{k})]$ and $[\bar{\pi}(2\hat{k}-n), \bar{\pi}(2(\hat{k}-1)-n)]$, and non-decreasing in π on each intervals $(\bar{\pi}(\hat{k}), \bar{\pi}(2\hat{k}-n))$ and $(\bar{\pi}(2\hat{k}-n), .5]$. By comparing across the intervals, we can also check that σ_g and σ_i are non-decreasing in π over $[0, .5]$ (Item 2).

For a level of belief π, the conviction probabilities of the guilty and the innocent, $\{(P_G, P_I)|\pi\}$, are determined by

$$P_G = \sum_{k'=\hat{k}}^{n} \binom{n}{k'} r_G^{k'}(1-r_G)^{n-k'}$$
$$P_I = \sum_{k'=\hat{k}}^{n} \binom{n}{k'} r_I^{k'}(1-r_I)^{n-k'}$$

where $r_G = p\sigma_g + (1-p)\sigma_i$ and $r_I = (1-p)\sigma_g + p\sigma_i$, where (σ_g, σ_i) are the efficient voting behavior. Note that the conviction probabilities of the guilty and the innocent are strictly increasing in r_G and r_I, which are strictly increasing in σ_g and σ_i.

Under equilibrium voting, $P_G \geq P_I$ clearly holds when the efficient voting behavior is $(\sigma_g = 0, \sigma_i = 0)$, since the conviction probabilities are all equal to zero. If the efficient voting behavior is responsive, we showed that (16) holds and $\frac{q}{1-q}\frac{1-\pi}{\pi} \geq 1$. Thus, $P_G \geq P_I$ (Item 3).

Moreover, $f_G(\pi)$ and $f_I(\pi)$ are non-decreasing in π, because conviction probabilities are strictly increasing in σ_g and σ_i, and σ_g and σ_i are non-decreasing in π (Item 4).

In addition to the properties in Proposition 1, we obtain the following lemma which is useful when we later prove Proposition 2. [19]

[18] The only special case is when the voting rule requires unanimity and $\pi = \bar{\pi}(n)$. Then any voting behavior of $(0 \leq \sigma_g \leq 1, \sigma_i = 0)$ is efficient.

[19] The lemma also holds for $f_I(\pi)$, but we do not need this observation in proving Proposition 2.

Lemma 5 *Conviction probability of the guilty $f_G(\pi)$ is an upper hemicontinuous correspondence in π with non-empty convex values.*

Proof: Note that the efficient voting behavior σ_g and σ_i are unique for every π, except when $\pi = \bar{\pi}(n)$ and the unanimity hold, where $\sigma_i = 0$ and σ_g can be any in $[0, 1]$. Since $\sum_{k'=\hat{k}}^{n} \binom{n}{k'} r_G^{k'}(1 - r_G)^{n-k'}$ is continuous in σ_g and σ_i, $f_G(\pi)$ is a convex valued for all π. In addition, closed form solutions of efficient voting behavior (σ_g and σ_i) are upper hemicontinuous in π. Since f_G is continuous in σ_g and σ_i, $f_G(\pi)$ inherits upper hemicontinuity in π. ∎

Appendix C Proof of Proposition 2

Suppose we have $\theta > P_G$ under an equilibrium. No defendant pleads guilty, and the jurors' reasonable beliefs π will be equal to .5. The conviction probabilities (P_G, P_I) must be in $\{(P'_G, P'_I)|.5\}$ (Item 1).

Otherwise, $\theta \leq P_G$. Note that $\theta \in [0, \bar{\theta}]$ where $\bar{\theta} := \sup f_G(.5)$. There exists a π such that $\theta = P_G \in f_G(\pi)$, because $f_G(\pi)$ is upper hemicontinuous in π with non-empty convex values (Intermediate Value Theorem). Suppose by a contradiction that $\theta < P_G$. Every guilty defendant pleads guilty, and only the innocent may or may not go to trial. In such case, jurors reasonably believe that all defendants in trials are innocent ($\pi = 0$), which consequently leads conviction probability equals to zero. This contradicts $\theta < P_G$. $\theta = P_G$ must be true (Item 2).

Appendix D Proof of Proposition 3

The prosecutor's problem is described below.

$$\max_{\theta\in[0,1]} -\frac{1}{2}q'\Big(\phi_I\theta+(1-\phi_I)P_I\Big)-\frac{1}{2}(1-q')\Big(\phi_G(1-\theta)+(1-\phi_G)(1-P_G)\Big) \tag{21}$$

$$\text{such that}\quad \begin{array}{ll} (a.1) & \phi_G \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_G \\ (a.2) & \phi_I \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_I \\ (b) & \pi = \begin{cases} 0 & \text{if } \phi_G=\phi_I=1 \\ \frac{1-\phi_G}{(1-\phi_G)+(1-\phi_I)} & \text{otherwise.} \end{cases} \\ (c) & (P_G,P_I)\in\{(P'_C,P'_I)|\pi\}. \end{array}$$

Using Proposition 2, we simplify the above expressions. To begin with, we can restrict without a loss of generality that a prosecutor can offer $\theta\in[0,\bar{\theta}]$, because he can obtain any utility level from offering $\theta>\bar{\theta}$ by offering $\theta=\bar{\theta}$: all players perceive the same ex-ante punishment in both cases. In the former case, all defendants plead not guilty and receive $(P_G,P_I)\in\{(P'_G,P'_I)|\pi\}$ conviction probabilities. In the latter case, some guilty defendants may plead guilty, but the punishment for a guilty plea is equal to conviction probability, which is the expected punishment from a jury trial. As far as the ex-ante punishments are same, the prosecutor and the defendant are indifferent pleading guilty and undergoing jury trial.

Once the prosecutor offers $\theta\in[0,\bar{\theta}]$, the second item in Proposition 2 ensures that $\theta=P_G\geq P_I$. Pleading decisions of the guilty are straightforward: the guilty are indifferent toward pleading guilty or pleading not guilty, thus any $\phi_G\in[0,1]$ is rational. Pleading decisions of the innocent depend on θ. $P_G=P_I$ holds only when $\theta=P_G=P_I=0$; otherwise, $\theta=P_G>P_I$. In the former case, any pleading decision behavior incurs the same expected prosecutor's utility, $-\frac{1}{2}(1-q')$, especially when $\phi_I=1$ (no punishment). In the latter case, $\phi_I=1$ must be true, since only pleading not guilty is rational. In all, when the prosecutor offers θ from $[0,\bar{\theta}]$, it is innocuous for the prosecutor to assume that $\phi_I=1$. By applying these observations, we simplify the prosecutor's decision as

$$\max_{\theta\in[0,\bar{\theta}]} -\frac{1}{2}q'P_I-\frac{1}{2}(1-q')(1-\theta)$$

$$\text{such that}\quad \begin{array}{ll} (a) & \phi_G\in[0,1] \\ (b) & \pi = \begin{cases} 0 & \text{if } \phi_G=1 \\ \frac{1-\phi_G}{2-\phi_G} & \text{otherwise.} \end{cases} \\ (c) & (\theta,P_I)\in\{(P'_G,P'_I)|\pi\}. \end{array}$$

We define a function $\tilde{P}_I : [0, \bar{\theta}] \to [0, 1]$ as follows

$$\tilde{P}_I(\theta) = p_I, \quad \text{where} \quad \exists\, \pi, \quad (\theta, p_I) \in \{(P'_G, P'_I) | \pi\}.$$

With referencing the proof of Proposition 1, we verify weather the function $\tilde{P}_I$ is well-defined: if the value of $\tilde{P}_I$ exists and is unique for all $\pi \in [0, \bar{\theta}]$. There are four cases: (1) $\theta = 0$, (2) $\theta \in (0, \hat{p_G})$, (3) $\theta = \hat{p_G}$, or (4) $\theta \in (\hat{p_G}, \bar{\theta}]$, where $\hat{p_G}$ is the conviction probability of the guilty when jurors vote by following their own signals ($\sigma_g = 1, \sigma_i = 0$).

If $\theta = 0$, p_I must be 0. If $\theta = \hat{p_G}$, p_I is unique and the value is derived from the voting strategy ($\sigma_g = 1, \sigma_i = 0$). For other cases, recall that the conviction probabilities are defined as

$$P_G = \sum_{k=\hat{k}}^{n} \binom{n}{k} r_G^k \big(1 - r_G\big)^{n-k}, \quad P_I = \sum_{k=\hat{k}}^{n} \binom{n}{k} r_I^k \big(1 - r_I\big)^{n-k}$$

where $r_G = p\sigma_g + (1-p)\sigma_i$ and $r_I = (1-p)\sigma_g + p\sigma_i$. When $\theta \in (0, \hat{p_G})$, $\sigma_i = 0$ and both P_G and P_I are strictly increasing in σ_g. Since P_G is continuous in r_G which is also continuous in σ_g, for any $\theta \in (0, \hat{p_G})$, there exists a unique σ_g inducing $\theta = P_G$. Such σ_g combined with $\sigma_i = 0$ gives a unique p_I such that $(\theta, p_I) \in \{(P'_G, P'_I)|\pi\}$. A similar procedure applies when $\theta \in (\hat{p_G}, \bar{\theta}]$.

Through the above argument, the function $\tilde{P}_I$ is not only well-defined, but strictly increasing and continuous on $[0, \bar{\theta}]$, and differentiable on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$. Using $\tilde{P}_I$, the prosecutor's problem becomes

$$\max_{\theta \in [0, \bar{\theta}]} U(\theta) := -\frac{1}{2} q' \tilde{P}_I(\theta) - \frac{1}{2}(1 - q')(1 - \theta). \tag{22}$$

We show that the objective function above is concave. Since $\tilde{P}_I$ is continuous in θ, the objective function is, too. Moreover, $\tilde{P}_I$ is differentiable on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$, and $U(\theta)$ is a linear combination of θ and $\tilde{P}_I$. Thus, $U(\theta)$ is also differentiable with respect to θ on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$. If we show that derivatives of $\tilde{P}_I$ is decreasing on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$, and the left derivate is greater than the right at $\hat{p_G}$, then the concavity of $\tilde{P}_I$ follows. Since $U(\theta)$ is a linear combination of θ and $\tilde{P}_I$, concavity of the objective function directly follows the concavity of $\tilde{P}_I$.

When $\theta \in (0, \hat{p_G})$, P_G and P_I are differentiable with respect to σ_g. The derivative of P_G is

$$
\begin{aligned}
\frac{\partial P_G}{\partial \sigma_g} &= \frac{\partial}{\partial \sigma_g} \sum_{k=\hat{k}}^{n} \binom{n}{k} (r_G)^k (1-r_G)^{n-k} \\
&= \sum_{k=\hat{k}}^{n-1} \Bigg(\frac{n!}{k!(n-k)!} k r_G^{k-1} (1-r_G)^{n-k} r'_G \\
&\quad - \frac{n!}{k!(n-k-1)!} r_G^k (n-k)(1-r_G)^{n-k-1} r'_G \Bigg) + n r_G^{n-1} r'_G \\
&= n\, r'_G \binom{n-1}{\hat{k}-1} r_G^{\hat{k}-1} (1-r_G)^{n-\hat{k}} \qquad (23)
\end{aligned}
$$

Using a similar operation, we obtain

$$
\frac{\partial P_I}{\partial \sigma_g} = n\, r'_I \binom{n-1}{\hat{k}-1} r_I^{\hat{k}-1} (1-r_I)^{n-\hat{k}} \tag{24}
$$

Therefore,

$$
\frac{\partial \tilde{P}_I(\theta)}{\partial \theta} = \frac{\partial P_I / \partial \sigma_g}{\partial P_G / \partial \sigma_g} = \frac{r'_I\, r_I^{\hat{k}-1} (1-r_I)^{n-\hat{k}}}{r'_G\, r_G^{\hat{k}-1} (1-r_G)^{n-\hat{k}}}. \tag{25}
$$

Since $r_G = p\sigma_g$ and $r_I = (1-p)\sigma_g$, (25) becomes

$$
\left(\frac{1-p}{p}\right)^{\hat{k}} \left(\frac{1-(1-p)\sigma_g}{1-p\sigma_g}\right)^{n-\hat{k}}. \tag{26}
$$

As θ increases in $(0, \hat{p_G})$, the corresponding σ_g increases, and the above derivative strictly decreases. Therefore, $\frac{\partial \tilde{P}_I(\theta)}{\partial \theta}$ is decreasing in $\theta \in (0, \hat{p_G})$.

When $\theta \in (\hat{p_G}, \bar{\theta})$, σ_g is fixed equal to 1 and only σ_i varies. Similar to (23) and (24), we obtain

$$
\frac{\partial \tilde{P}_I(\theta)}{\partial \theta} = \frac{\partial P_I / \partial \sigma_i}{\partial P_G / \partial \sigma_i} = \frac{r'_I\, r_I^{\hat{k}-1} (1-r_I)^{n-\hat{k}}}{r'_G\, r_G^{\hat{k}-1} (1-r_G)^{n-\hat{k}}}. \tag{27}
$$

By substituting in $r_G = p + (1-p)\sigma_i$ and $r_I = (1-p) + p\sigma_i$, we obtain

$$\left(\frac{(1-p)+p\sigma_i}{p+(1-p)\sigma_i}\right)^{\hat{k}-1}\left(\frac{p}{1-p}\right)^{n-\hat{k}+1}. \tag{28}$$

Again, as θ increases in $(\hat{p_G}, \bar{\theta})$, the corresponding σ_i increases, and the above derivative decreases. Therefore, $\frac{\partial \tilde{P}_I(\theta)}{\partial \theta}$ is decreasing in θ.

Lastly, at $\theta = \hat{p_G}$, the left derivative is greater than the right derivative, because the limit of (26) as σ_g goes to 1 is greater than the limit of (28) as σ_i goes to 0. This concludes that $\tilde{P}_I$ is strictly concave in θ, and thus the objective function in (22) is also strictly concave in θ.

Since the prosecutor's objective function is strictly concave in θ, First Order Condition (FOC) gives the necessary and sufficient condition of optimizer θ^*. Proposition 3 also narrates the first order condition. Instead of finding the closed form solution, we study the FOC. Provided that the optimizer θ^* falls into,

Interior Solutions

$[0 < \theta^* < \hat{p_G}]$: Using (26), FOC of (22) becomes

$$\left(\frac{p}{1-p}\right)^{\hat{k}}\left(\frac{1-p\sigma_g}{1-(1-p)\sigma_g}\right)^{n-\hat{k}} = \frac{q'}{1-q'}.$$

Recall that a juror receiving guilty signal uses a mixed strategy in this level of conviction probability for the guilty. (Equation (12) holds.) We obtain

$$\frac{q}{1-q}\frac{1-\pi}{\pi} = \frac{q'}{1-q'}$$

$[\hat{p_G} < \theta^* < \bar{\theta}]$: Using (28), FOC of (22) becomes

$$\left(\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}\right)^{\hat{k}-1}\left(\frac{1-p}{p}\right)^{n-\hat{k}+1} = \frac{q'}{1-q'}.$$

Recall that a juror receiving innocent signal uses a mixed strategy in this level of conviction probability for the guilty. (Equation (15) holds.) We obtain

$$\frac{q}{1-q}\frac{1-\pi}{\pi} = \frac{q'}{1-q'}$$

Boundary Solutions

$[\theta^* = \hat{p_G}]$: The prosecutor offer such punishment for a guilty plea, when

$$\lim_{\theta \downarrow \hat{p_G}} \frac{\partial U(\theta)}{\partial \theta} \leq 0 \leq \lim_{\theta \uparrow \hat{p_G}} \frac{\partial U(\theta)}{\partial \theta}$$

Replacing (26) and (28) for $\frac{\partial \tilde{P}_I(\theta)}{\partial \theta}$, we can rewrite the above inequalities as

$$\left(\frac{(1-p)+p\sigma_i}{p+(1-p)\sigma_i}\right)^{\hat{k}-1}\left(\frac{p}{1-p}\right)^{n-\hat{k}+1} \leq \frac{1-q'}{q'} \leq \left(\frac{1-p}{p}\right)^{\hat{k}}\left(\frac{1-(1-p)\sigma_g}{1-p\sigma_g}\right)^{n-\hat{k}},$$

or

$$\left(\frac{p}{1-p}\right)^{2(\hat{k}-1)-n} \leq \frac{q'}{1-q'} \leq \left(\frac{p}{1-p}\right)^{2\hat{k}-n}$$

Compared with (13), when the prosecutor chooses $\theta^* = \hat{p_G}$, the jurors' voting behavior with π and q is exactly the same as the voting behavior when jurors' beliefs are equal to .5 and reasonable doubt are equal to q'.

$[\theta^* = 0]$: The right derivative at $\theta = 0$ must be non-positive. Applying (26) to the derivative of the objective function in (22) while taking $\sigma_g \to 0$, we obtain

$$\left(\frac{p}{1-p}\right)^{\hat{k}} \leq \frac{q'}{1-q'}.$$

Note that such θ^* induces the equilibrium voting behavior $\sigma_g = \sigma_i = 0$. This strategy profile becomes an efficient voting behavior when the RHS of (11) is greater than or equal to the LHS, which implies

$$\left(\frac{p}{1-p}\right)^{\hat{k}} \frac{\pi}{1-\pi} \leq \frac{q}{1-q}.$$

By comparing the above two inequalities, we observe that the equilibrium voting behavior is the same as the voting behavior when jurors' beliefs are equal to .5 and reasonable doubts are equal to q'.

[$\theta^* = \bar{\theta}$]: The left derivative at $\theta = \bar{\theta}$ must be non-negative. Applying (28) to the derivative of $U(\theta)$, we must obtain

$$\lim_{\theta \uparrow \bar{\theta}} \frac{\partial U(\theta)}{\partial \theta} \geq 0$$

or

$$\left(\frac{p + (1-p)\bar{\sigma_i}}{(1-p) + p\bar{\sigma_i}}\right)^{\hat{k}-1} \left(\frac{1-p}{p}\right)^{n-\hat{k}+1} \geq \frac{q'}{1-q'}$$

where $\bar{\sigma_i}$ with $\sigma_g = 1$ is an equilibrium voting behavior with the belief $\pi = .5$.

Note that in this situation, a juror receiving an innocent signal is indifferent between conviction and acquittal. Thus (14) becomes

$$\left(\frac{p + (1-p)\bar{\sigma_i}}{(1-p) + p\bar{\sigma_i}}\right)^{\hat{k}-1} \left(\frac{1-p}{p}\right)^{n-\hat{k}+1} = \frac{q}{1-q}.$$

Thus, $\frac{q}{1-q} \geq \frac{q'}{1-q'}$, or $q \geq q'$.

When $q \geq q'$, the prosecutor offers $\theta^* = \bar{\theta}$, and all defendants plead not guilty ($\pi = .5$). Jurors vote with threshold $\frac{q}{1-q}$, which is the same as the threshold in jury model without plea bargaining. Although we have restricted the prosecutor's strategy space to $[0, \bar{\theta}]$, any θ^* higher than $\bar{\theta}$ induces the same prosecutor's equilibrium expected utility as $\theta^* = \bar{\theta}$.

Appendix E Proof of Corollary 4

First, note that efficient voting behavior is responsive if $\pi > \bar{\pi}(\hat{k})$. Since $\bar{\pi}(l)$ is strictly decreasing in l, the efficient voting behaviors are responsive for all $\pi > 0$ as $n \to \infty$.

Given π, p, and unanimity ($\hat{k} = n$), the efficient voting leads the conviction probabilities to converge to $1 - \left(\frac{(1-q)(1-p)\pi}{qp(1-\pi)}\right)^{1-\frac{p}{2p-1}}$ for the guilty, and to $\left(\frac{(1-q)(1-p)\pi}{qp(1-\pi)}\right)^{\frac{p}{2p-1}}$ for the innocent. These convergence results directly follow Proposition 2 in Feddersen and Pesendorfer (1998). (Our parameter values satisfy all conditions assumed in their Proposition.)

For non-unanimous rules, regardless of the jury size n, we have $\frac{\pi}{1-\pi} = 1$ (if $q > q'$) or $\frac{1-q}{q}\frac{\pi}{1-\pi} = \frac{1-q'}{q'}$ (if $q \leq q'$). As we replace $\frac{1-q}{q}\frac{\pi}{1-\pi} = \frac{1-\tilde{q}}{\tilde{q}}$ where $\tilde{q} = \max\{q, q'\}$, the conviction probabilities for the guilty and the innocent directly follow Proposition 3 in Feddersen and Pesendorfer (1998): conviction probability for the guilty converges to 1 and for the innocent converges to 0.

Lastly from Proposition 2, we can relate the ex-ante punishments, one for the guilty and another for the innocent, to the conviction probabilities in jury trials.

References

Austen-Smith, D., and J. S. Banks (1996): "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *The American Political Science Review*, 90(1), 34–45.

Austen-Smith, D., and T. Feddersen (2003): "Deliberation and voting rules," *Northwestern University Typescript*.

Bibas, S. (2004): "Plea Bargaining outside the Shadow of Trial," *Harvard Law Review*, 117(8), 2463–2547.

Cho, I.-K., and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics*, 102(2), 179–221.

Condorcet, M. (1785): "Essai sur lapplication de lanalyse à la probabilité des decisions rendues a la pluralité des voix," *Paris: Limprimerie royale*.

Cooter, R., and D. Rubinfeld (1989): "Economic analysis of legal disputes and their resolution," *Journal of Economic Literature*, 27(3), 1067–1097.

Coughlan, P. (2000): "In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting," *The American Political Science Review*, 94(2), 375–393.

Feddersen, T., and W. Pesendorfer (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *The American Political Science Review*, 92(1), 23–35.

Feddersen, T. J., and W. Pesendorfer (1996): "The Swing Voter's Curse," *The American Economic Review*, 86(3), 408–424.

Gerardi, D., and L. Yariv (2007): "Deliberative voting," *Journal of Economic Theory*, 134(1), 317–338.

GOEREE, J., AND L. YARIV (2010): "An experimental study of jury deliberation," .

GROSSMAN, G., AND M. KATZ (1983): "Plea bargaining and social welfare," *The American Economic Review*, 73(4), 749–757.

GUARNASCHELLI, S., R. D. MCKELVEY, AND T. R. PALFREY (2000): "An Experimental Study of Jury Decision Rules," *The American Political Science Review*, 94(2), 407–423.

MNOOKIN, R. H., AND L. KORNHAUSER (1979): "Bargaining in the Shadow of the Law: The Case of Divorce," *The Yale Law Journal*, 88(5), 950–997.

NASH, J. (1951): "Non-cooperative games," *Annals of mathematics*, 54(2), 286–295.

PRIEST, G. L., AND B. KLEIN (1984): "The Selection of Disputes for Litigation," *The Journal of Legal Studies*, 13(1), 1–55.

STUNTZ, W. J. (2004): "Plea Bargaining and Criminal Law's Disappearing Shadow," *Harvard Law Review*, 117(8), 2548–2569.