Plea Bargaining: On the Selection of Jury Trials

SangMok Lee*

January 2, 2014

Abstract

We consider a model of the criminal court process, focusing on plea bargaining. A plea bargain provides unequal incentives to go to trial because innocent defendants are more willing to plead not guilty. We show that the court process implements the preferences of the person or group who is most concerned about wrongful conviction. If a prosecutor is more concerned about wrongful conviction than the jury, the prosecutor can shape the defendant pool at trial so that jurors act according to prosecutor's preferences against judicial mistakes. Our model also connects insights from strategic jury models that usually omit plea bargaining with the actual criminal court process where most cases are resolved through plea bargaining. As an example, we show that the inferiority of the unanimity rule established in Feddersen and Pesendorfer (1998) persists in spite of the addition of plea bargaining.

JEL Classification Numbers: C72, D71, K41

Keywords: Plea bargaining, Strategic voting, Collective choice, Jury trial.

*Department of Economics, University of Pennsylvania, Philadelphia, PA, 19104. Email: sangmok-at-sas.upenn.edu. This paper is a part of my dissertation at Caltech. I am grateful to Leeat Yariv for encouragement and guidance. I also thank Luke Boosey, Kim Border, Brendan Daley, John Duggan, Federico Echenique, Matias Iaryczower, Navin Kartik, Morgan Kousser, Stephen Morris, Wojciech Olszewski, Jean-Laurent Rosenthal, Thomas Ruchti, Matthew Shum, Colin Stewart, and two anonymous referees for their helpful comments and suggestions.

1 Introduction

1.1 Overview

In the United States most criminal cases are resolved out of court by plea bargaining. The defendants often plead guilty in exchange for a more lenient sentence than the one they might receive with a jury trial. Among the 74,782 convictions in federal courts in 2004, 96% were achieved through plea bargaining. The rate increased from 87% in 1990 to 96% in 2004 for felony offenses.[1]

Formal economic analysis of plea bargaining has largely focused on the rather explicit welfare effects. Most evidently, plea bargains save trial costs for prosecutors, defendants, and jurors who are involved in the criminal court process (Rabe and Champion, 2002, p. 306-308). Plea bargains also give defendants a chance to avoid the risk of conviction at trial on more serious charges (Grossman and Katz, 1983).

In this paper, we focus on a less obvious but potentially crucial effect of plea bargaining: influence on jury behavior. In principle, the court instructs jurors to render a verdict according only to the presented evidence and to the instructions of the court: see, e.g., the California Code of Civil Procedure–Section 232 (b). However, jury trials are *chosen* by defendants who plead not guilty, and those defendants are more likely to be innocent. Therefore, it is natural to believe that only a small fraction of criminal cases going to jury trials may give jurors a first impression that the defendants are possibly innocent. We study such an effect of plea bargaining on jury behavior. In particular, the prosecutor can use plea bargaining to shape the defendant pool at trial, which influences jurors' beliefs about how likely a defendant at a trial is truly guilty. As a result, the prosecutor can bias jurors' voting behavior either for conviction or for acquittal.

The model is effectively a screening and signaling game. A plea bargain allows a prosecutor to *screen* out some defendants before going to trial. A defendant as a sender *signals* his type by pleading either guilty or not guilty, and jurors as receivers update their beliefs on the sender's type.[2] To emphasize the screening and signaling effect, we intentionally ignore all costs of going to a jury trial. Agents are also risk neutral: there is no cost of uncertainty from going to trial.

The model starts with a prosecutor indicting a defendant who is either guilty or innocent.

[1] See Table 4.2 in the Compendium of Federal Justice Statistics 2004; U.S. Department of Justice, Bureau of Justice Statistics: `http://bjs.ojp.usdoj.gov/content/pub/pdf/cfjs04.pdf`.

[2] We refer to prosecutors and defendants as male and jurors as female.

Without knowing the defendant's type, the prosecutor initiates a plea bargain offer. If the defendant pleads guilty, the case terminates with the offered punishment, or otherwise a jury trial follows. During testimony, each juror receives a private signal, which represents the influence of the trial evidence. The signal is imperfectly correlated with the defendant's type. If a super-majority of jurors vote for conviction (e.g., a two-thirds majority), the jury returns a guilty verdict. Otherwise, the jury returns a not-guilty verdict. The prosecutor and jurors have preferences against mistakenly delivered punishments to innocent defendants or undelivered punishments to guilty ones.[3]

We find that the court process implements the preferences of the person or group who is most concerned about wrongful conviction. If a prosecutor is more concerned about wrongful conviction than jurors, the prosecutor shapes the defendant pool at trial, which induces the jurors to vote as if they had the prosecutor's preferences. If the prosecutor is less concerned about wrongful conviction, the prosecutor finds no incentive to use plea bargaining, so all judicial decisions are made according to jurors' preferences. To understand how this occurs, consider the following lines of reasoning.

If a guilty defendant considers a plea bargain offer acceptable, he will plead guilty. Jurors will subsequently account for the decreased chance of having a guilty defendant at trial, which will lower the probability of conviction. This low conviction rate feeds back to plea bargaining. The previously acceptable offer becomes unacceptable for the guilty defendant, and the opposite story follows. A guilty defendant now considers the plea bargain offer unacceptable. This not guilty plea increases the chance of having a guilty defendant at trial, which raises the probability of conviction. In equilibrium, guilty defendants will be indifferent between taking a guilty plea punishment and undergoing a jury trial. Meanwhile, an innocent defendant chooses to go to a jury trial because he is less likely to be convicted.

The prosecutor attempts to devise a plea bargain offer that ultimately leads to his ideal conviction probabilities. Suppose the prosecutor cares more than the jurors about mistakenly delivering punishment to innocent defendants. If the prosecutor lowers the guilty plea charge, a guilty defendant is more likely to plead guilty, and a defendant in trial is more likely to be innocent. As a result, jurors are more careful to avoid convicting an innocent defendant. However, such influence is possible only in one direction: leading jurors to vote more frequently for acquittal. As guilty defendants are more likely to take plea bargain offers, plea bargains can only *decrease* the likelihood of a guilty defendant at trial. At least

[3]In this paper, the prosecutor may not single-mindedly pursue convictions. In practice, mismanaged cases may later become public knowledge, and such exposure will affect a prosecutor's future career. Even a self-interested prosecutor will be concerned about false prosecutions.

as a screening device, plea bargaining is of no use if the prosecutor cares less about convicting an innocent defendant. In this case, all judicial decisions are made according to jurors' preferences.

Our unified model of plea bargaining and a jury trial can extend the scope of strategic voting models, which often omit plea bargaining, to the actual court process, where most cases are resolved through plea bargaining. As an example, we revisit the comparison between the unanimity rule and general super-majority rules studied in Feddersen and Pesendorfer (1998). In a jury trial model without plea bargaining, with a moderate number of jurors, the probabilities of convicting an innocent defendant and acquitting a guilty defendant under the unanimity rule are significantly higher than those probabilities under any general super-majority rule. We show that the inferiority of the unanimity rule persists with the addition of plea bargaining. The expected value of punishment mistakenly delivered to innocent defendants or not delivered to guilty defendants under the unanimity rule are significantly higher than the expected value under any general super-majority rule.

Our model is flexible in regards to the choice of the jury trial model. The main idea, that plea bargaining leads to jurors' biased belief and voting behavior, remains valid for any reasonable model of jury trial. However, the inferiority of the unanimity rule may change. For example, Coughlan (2000) shows that the unanimity rule usually performs better than any super-majority rule if jurors can deliberate by pooling their private information. The unanimity rule induces a Nash equilibrium in which jurors honestly pool their private information and vote unanimously for a choice mutually better for all jurors. If we adopt Coughlan's model, the comparison between the unanimity and general super-majority rules would be reversed.

1.2 Related literature

The closest studies to this paper are Priest and Klein (1984) and Bjerk (2007). Priest and Klein consider a unified model of pre-trial processes and a jury trial to highlight that pools of defendants change over the court processes. In Bjerk (2007), a prosecutor receives an initial signal of a defendant's type and makes a plea bargain offer. If the defendant rejects the offer and goes to a jury trial, the jurors receive an updated signal and vote for conviction or acquittal. The main question is quite different from ours: the prosecutor and the jurors are assumed to have the same preferences, and the question is whether the prosecutor can induce their mutually optimal judicial outcomes by using plea bargains.

Studies on criminal court process often assume exogenously given jury behavior. Gross-

man and Katz (1983) study plea bargain as a screening device, which sorts guilty and innocent defendants through a self-selection mechanism. However, conviction probabilities at jury trial are fixed and never influenced by the plea bargain. Reinganum (1988) also focus on bargaining behavior, ignoring its effects on jury behavior. In her model, the prosecutor knows the strength of the case, and the defendant knows whether he is guilty or innocent. Lastly, in Baker and Mezzetti (2001), when a case goes to trial, the prosecutor further investigates the case, and only the additional information affects the jury behavior.

Plea bargaining has been studied predominantly as a *bargaining* problem. A jury trial costs time, effort, and uncertainty on final outcomes. Participants in a plea bargain can share a surplus by not going to jury trial, which is a typical bargaining problem. For a brief summary of this approach, see e.g., Cooter and Rubinfeld (1989). In this paper, we exclude count bargaining, in which defendants plead guilty to a subset of multiple original charges. For a model of bargaining over multiple issues, see, e.g., Busch and Horstmann (1999).

We adopt a strategic voting model as a benchmark model of jury behavior. Each juror receives private information during testimonies and votes based on her private information and the condition of being pivotal. In some scenarios, her pivotal position convinces her to follow other jurors' votes against her private information (Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1996). Feddersen and Pesendorfer (1998) apply this strategic voting behavior to a jury trial and find that the unanimity rule is inferior to any general super-majority rule.

The rest of this paper is organized as follows. We introduce our model in Section 2. In Section 3, we find equilibrium restrictions in jurors' voting behavior. The results in this section also serve as benchmark results of a jury trial without plea bargaining. In Section 4, we find equilibrium restrictions in plea bargaining. As an application of our model, we revisit the inferiority of the unanimity rule in Section 5. The conclusion of the paper is provided in Section 6. All proofs are relegated to the Appendix.

2 The Model

There are three types of agents: a prosecutor, a defendant, and jurors. The punishment after being convicted in a jury trial is normalized to 1. We assume that the defendant is guilty (G) with probability π_0, and otherwise is innocent (I). Only the defendant knows whether he is either guilty (G) or innocent (I).

A criminal court process consists of two phases:

- $t = 1$**: A plea bargain occurs.**

  The prosecutor makes a plea bargain offer $\theta \in [0, 1]$. The defendant pleads either *guilty* or *not guilty*. If the defendant pleads guilty, the case terminates and the punishment θ is delivered. Otherwise, the plea bargain is withdrawn, and the case proceeds to the second phase described below.[4]

- $t = 2$**: A jury trial occurs.**

  A jury consists of n ($n > 1$) jurors and a voting rule $\hat{k}$ ($1 < \hat{k} \leq n$). Each juror receives a private signal g or i, which is positively correlated with the true state G or I, as given by

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

  We assume that $p \in (.5, 1)$: each juror receives a correct signal with probability p, and receives an incorrect signal with probability $1 - p$.[5]

  The jury reaches a decision by casting votes simultaneously. If the number of vote to convict is larger than or equal to the voting rule $\hat{k}$, the defendant is convicted (C). Otherwise, the defendant is acquitted (A). We call a rule requiring $\hat{k} = n$ votes for conviction *the unanimity rule*, and others *general super-majority rules*.

Each type of agent has a utility function defined as follows:

- **A defendant:**

  Utilities are determined negatively by the amount of punishment: -1 if he is convicted, 0 if he is acquitted, and $-\theta$ if he pleads guilty. The defendant is assumed to be risk neutral: if a defendant perceives that he will be convicted with probability h, the expected utility of going to trial is $-h = h \cdot (-1) + (1 - h) \cdot 0$, which is the same utility from deterministic punishment h.

- **Jurors:**

[4] We consider a simple plea bargaining setup to highlight how the plea bargaining affects jurors' voting behavior. A more natural model of plea bargaining would be a dynamic setup with asymmetric information where both participants can make offers. See, e.g., Inderst (2003) for a model of bargaining with one-sided private information and alternating offers. In the context of buyer-seller bargaining over a divisible good, Bac (2000) shows that the participant with informational advantage (cf, a defendant in plea bargaining) may strategically delay offers which are restricted to only a portion of the good (cf, a part of criminal charges).

[5] Each juror may have a different interpretation during the testimony by witnesses due to her personal background. The private signal captures such interpretation.

Jurors' utility from correct judicial decisions is normalized to $u[C|G] = u[A|I] = 0$. Convicting innocent defendants or acquitting guilty defendants incurs utility losses, $u[C|I] = -q$ and $u[A|G] = -(1-q)$, respectively. We assume that $q \in [.5, 1)$.

We term q as *the threshold level of reasonable doubt* following Feddersen and Pesendorfer (1998). Suppose a juror, with all the information available to her, believes that a defendant is guilty with probability $\tilde{q}$. The expected utility from conviction $-q(1-\tilde{q})$ is greater than or equal to the expected utility from acquittal $-(1-q)\tilde{q}$ if and only if $\tilde{q} \geq q$. Therefore, jurors use q as the threshold level of belief to vote for conviction.

- **A prosecutor:**

  When punishment $h \in [0, 1]$ is delivered to a defendant, the prosecutor's utility is given by

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

  where $q' \in [0, 1]$.

  The prosecutor loses utility when punishments are delivered to innocent defendants, or guilty defendants avoid their just punishments. The prosecutor is assumed to be risk neutral: he is indifferent between delivering h punishment through plea bargaining and going to trial where the trial convicts the defendant with probability h.

We denote by ϕ_G the probability that a guilty defendant pleads guilty, and by ϕ_I the probability that an innocent defendant pleads guilty. These probabilities are strategies for defendants that will arise in equilibrium; they are not being exogenously imposed. The jurors have a common posterior belief $\pi \in [0, 1]$ that a defendant is guilty given that the case comes to a jury trial. Each juror j makes her voting decision: voting for conviction with probability σ_g^j when her signal is g and with probability σ_i^j if the signal is i. We assume that jurors do not observe the terms of plea offer declined by the defendant.[6]

Figure 1 summarizes the timing of the model. A prosecutor offers θ in a plea bargain, and a defendant pleads either guilty or not guilty. If the defendant pleads guilty, the case terminates with delivering θ punishment to the defendant. If the defendant pleads not guilty, the case goes to a jury trial. The jury determines whether to convict or acquit.

We find a Perfect Bayesian Equilibrium with two equilibrium refinements: one in jurors' voting behavior and the other in jurors' common beliefs.

[6]In fact, the equilibrium does not change if we assume that jurors observe the terms of declined plea offer. Given a common posterior belief π, the jurors need not take into account the terms of plea offer in their voting behavior.

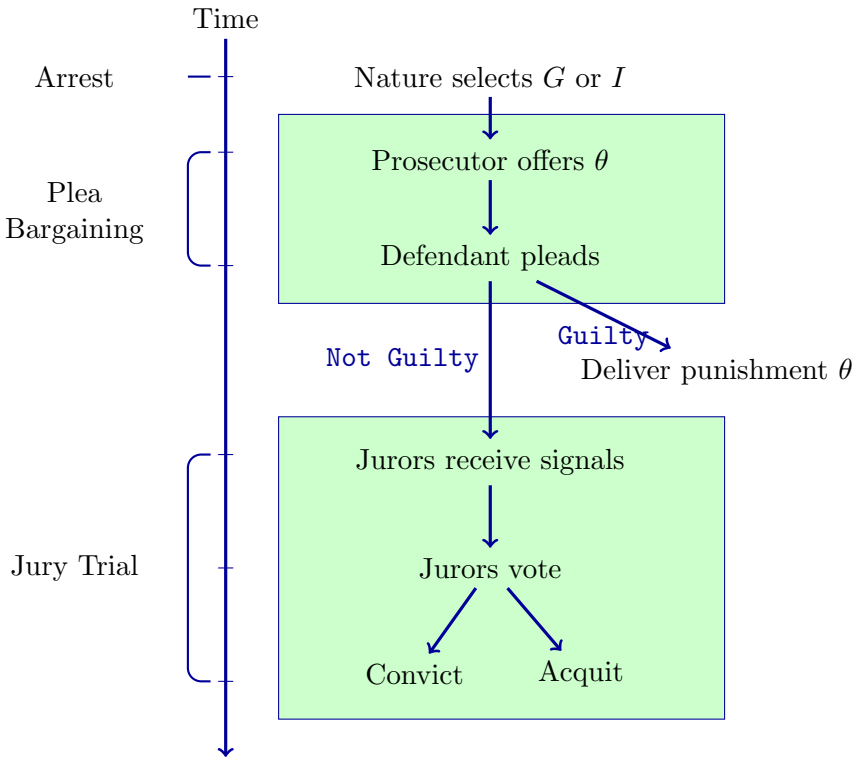


Figure 1: A criminal court process.

For jurors' voting behavior, we consider *symmetric equilibrium voting behavior*: all jurors adopt the same strategy. We denote a symmetric strategy profile by (σ_g, σ_i) without specifying a particular juror. A symmetric voting behavior is called *responsive* if the probability for a juror to vote for conviction with signal g is strictly higher than the probability with signal i (i.e., $\sigma_g > \sigma_i$). We find symmetric voting behavior that gives jurors the highest expected payoffs. We call this refined equilibrium behavior *the most efficient symmetric equilibrium voting behavior* or, more succinctly, *the efficient equilibrium voting behavior*. Whenever a responsive symmetric equilibrium exists, the most efficient equilibrium rules out trivial voting behavior (e.g., voting always for acquittal). Trivial voting behavior forms a symmetric equilibrium because no juror is pivotal, so every juror is indifferent between voting for conviction and voting for acquittal. Clearly, a trivial voting equilibrium is inefficient as jurors do not use their private information in their voting decisions.[7]

Jurors' beliefs also need a refinement when the probability of a defendant going to trial is zero. We refine jurors' beliefs that a defendant coming to trial is believed to be innocent. Such refinement is equivalent to imposing D1 by Cho and Kreps (1987) on the equilibria of the signaling game, which is induced from our model by assuming that the jurors follow the efficient equilibrium voting behavior.

In the spirit of backward induction, we first study jurors' efficient equilibrium voting behavior, and then equilibrium behaviors in plea bargaining. The following section on jury trial is a part of the backward induction. The section also serves as a baseline for comparison as a jury model without plea bargaining.

3 A Jury Trial

The jurors' have a common belief: conditional on a case coming to trial, the defendant is guilty with probability π. Note that a guilty defendant has a higher chance of being convicted because he is more likely to generate guilty signals than an innocent defendant and each juror is more likely to vote for conviction when her signal is g. Therefore, we assume without loss of generality that a guilty defendant is more likely to plead guilty ($\phi_G \geq \phi_I$) and less likely to come to trial ($\pi \leq \pi_0$).

Following standard strategic jury models, we assume that a juror makes her voting decision under the condition that she is pivotal (piv). We denote by $P[G|piv, g, \pi]$ the posterior

[7]Other notions of refinement motivated by the *trembling hand perfection* in Austen-Smith and Feddersen (2005) or *weakly undominated strategies* in Gerardi and Yariv (2007) are insufficient to get a unique equilibrium (see Appendix E).

probability that the defendant turns out to be guilty when a juror receives a guilty signal g, has a belief π, and is pivotal:

$$Pr[G|piv, g, \pi] := \frac{Pr[piv|G] \cdot p \cdot \pi}{Pr[piv|G] \cdot p \cdot \pi + Pr[piv|I] \cdot (1-p) \cdot (1-\pi)}. \tag{2}$$

Her expected utility from a guilty verdict is $-q \cdot Pr[I|piv, g, \pi]$ and the utility from a not guilty verdict is $-(1-q) \cdot Pr[G|piv, g, \pi]$. Given all the information available to the juror, she will vote for conviction if

$$-q \cdot Pr[I|piv, g, \pi] \geq -(1-q) \cdot Pr[G|piv, g, \pi],$$

or equivalently,

$$Pr[G|piv, g, \pi] \geq q.$$

That is, the evidence of guilt is enough to exceed the level of reasonable doubt q.

Thus, a necessary condition for a juror who receives a guilty signal to vote for conviction (or acquittal) is

$$Pr[G|piv, g, \pi] \geq \text{ (or } \leq) \; q. \tag{3}$$

If the inequality holds as equality, the juror may vote for conviction or acquittal with certain probabilities.

We find an equilibrium voting behavior by explicitly computing $Pr[G|piv, g, \pi]$. For now, assume that $Pr[piv|G]$, $Pr[piv|I]$, and π are all strictly positive. We rearrange Equation (2) that

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

Combined with Equation (3), a juror with a guilty signal will vote for conviction (or acquittal) when

$$\frac{Pr[piv|G]}{Pr[piv|I]} \frac{p}{1-p} \frac{\pi}{1-\pi} \geq \text{ (or } \leq) \; \frac{q}{1-q}.$$

The left-hand side is the likelihood ratio of guilty to innocent given that a juror is pivotal, multiplied by the likelihood ratio inferred from private information g, multiplied by the ratio of beliefs on the defendant's type. The right-hand side is the ratio of the two kinds of utility loses.

Let r_G denote the probability that a juror votes for conviction when the defendant is truly

guilty, and let r_I be the same probability when the defendant is instead innocent. That is,

$$r_G := p\sigma_g + (1-p)\sigma_i \quad \text{and} \quad r_I := (1-p)\sigma_g + p\sigma_i.$$

If a voting rule requires $\hat{k}$ $(1 < \hat{k} \leq n)$ number of conviction votes for a guilty verdict, a juror is pivotal when $\hat{k}-1$ other jurors vote for conviction. Thus, when a juror with a guilty signal votes for conviction (or acquittal), we must have

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{p}{1-p} \frac{\pi}{1-\pi} \quad \geq (\text{or} \leq) \quad \frac{q}{1-q}. \tag{4}$$

We obtain a similar voting criterion for a juror who receives an innocent signal:

$$\frac{r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}} \frac{1-p}{p} \frac{\pi}{1-\pi} \quad \geq (\text{or} \leq) \quad \frac{q}{1-q}. \tag{5}$$

The above expressions are the main restrictions of jurors' equilibrium behavior.

To understand how jurors' beliefs affect the equilibrium voting behavior, it is convenient to introduce a function $\bar{\pi}$ defined as

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

The motivation behind the definition of $\bar{\pi}$ becomes clear if we rearrange the equation:

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

The function $\bar{\pi}$ maps a number of guilty signals l to a level of belief π, which gives the minimum amount of evidence for a conviction vote. If a jury consists of a single juror who receives l number of independent signals, $\bar{\pi}(l)$ is the threshold level of the juror's belief such that the juror votes for conviction if all l signals are guilty.

We state the equilibrium voting behavior in the following proposition.

Proposition 1 (Equilibrium Voting Behavior) *If $\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(-n+\hat{k}-1)$, the most efficient symmetric equilibrium voting behavior is responsive. Otherwise, the most efficient symmetric equilibrium involves an equilibrium in which no juror votes for conviction (if $\pi \leq \bar{\pi}(\hat{k})$), or all jurors vote for conviction (if $\pi \geq \bar{\pi}(-n+\hat{k}-1)$).*

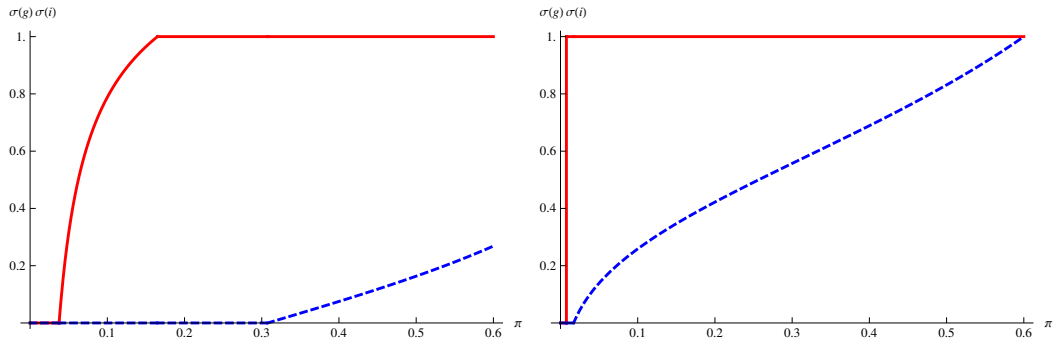
When a responsive equilibrium voting behavior exists, Proposition 1 rules out trivial equilibria where jurors always vote for conviction or always vote for acquittal. In some situations, such voting behavior is an equilibrium behavior because no juror becomes pivotal. A juror's vote never changes the judicial decision, so every juror is indifferent between voting for conviction and voting for acquittal. Intuitively, if a responsive equilibrium exists, it must be more efficient than the trivial equilibria because jurors *use* their signals in voting decisions.

The only exception is under the unanimity rule ($\hat{k} = n$) when $\pi = \bar{\pi}(n)$. In this situation, efficient equilibria involve both responsive equilibrium voting behavior and a trivial equilibrium behavior where all jurors vote for acquittal. The belief $\bar{\pi}(n)$ is so low that even with n guilty signals each juror is indifferent between a guilty and an innocent verdict. Then, the responsive equilibrium is not necessarily more efficient than the trivial equilibrium of always voting for acquittal.

We illustrate how to find an equilibrium voting behavior from voting criteria (4) and (5). Suppose jurors do not always vote for acquittal ($0 < r_I, r_G$) and do not always vote for conviction ($r_I, r_G < 1$). Then the left-hand side of (4) is strictly larger than the left-hand side of (5). Thus, a juror with signal g has a greater probability of voting for conviction than a juror with signal i: i.e., $\sigma_g > \sigma_i$. Three classes of strategies are consistent with such jury behavior: ($0 < \sigma_g < 1, \sigma_i = 0$), ($\sigma_g = 1, 0 < \sigma_i < 1$), and ($\sigma_g = 1, \sigma_i = 0$).

Given a voting rule requiring $\hat{k}$, ($\sigma_g = 1, \sigma_i = 0$) is not an equilibrium behavior for $\pi < \bar{\pi}(2\hat{k} - n)$. Suppose that a juror with signal g turns out to be pivotal. That is, $\hat{k} - 1$ other jurors vote for conviction and $n - \hat{k}$ jurors vote for acquittal. Considering that other 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. Since signals are symmetric ($P[g|G] = P[i|I]$), being pivotal is then equivalent to observing $2\hat{k} - n - 1$ guilty signals, which results in $2\hat{k} - n$ guilty signals, including the juror's own. When $\pi < \bar{\pi}(2\hat{k} - n)$, the $2\hat{k} - n$ guilty signals provide insufficient evidence of a guilty verdict. Thus, $\sigma_g = 1$ is not a best response, and ($\sigma_g = 1, \sigma_i = 0$) must not be an equilibrium voting behavior.

When a juror with signal g uses a mixed strategy ($0 < \sigma_g < 1, \sigma_i = 0$), she must be indifferent between conviction and acquittal. In such an instance, the voting criterion (4) holds with an equality, from which we obtain an expression for σ_g and the consistent range of π. When a juror with a signal i uses a mixed strategy ($\sigma_g = 1, 0 < \sigma_i < 1$), we obtain σ_i and the range of π from the equality of voting criterion (5). If jurors with signal g vote for conviction and with signal i vote for acquittal ($\sigma_g = 1, \sigma_i = 0$), the juror with signal g has enough evidence to vote for conviction, whereas a juror with signal i lacks evidence and



(a) A super-majority 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 $q = \frac{1}{2}$

thus votes for acquittal. The corresponding inequalities of voting criteria (4) and (5) allow us to find the range of π consistent with such a strategy.

Figure 2 depicts efficient equilibrium voting behavior under a general super-majority rule ($1 < \hat{k} < n$) and the unanimity rule ($\hat{k} = n$). The solid lines represent the probabilities of voting for conviction with signal g, and the dashed lines represent the probabilities with signal i. Mostly, we have a unique equilibrium voting behavior, except when $\pi = \bar{\pi}(\hat{k})$ under the unanimity rule.

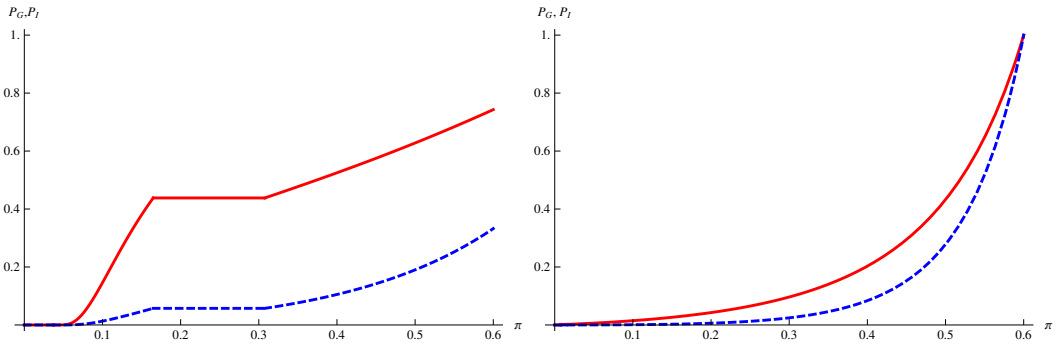
We next find equilibrium conviction probabilities P_G and P_I of a guilty defendant and an innocent defendant, respectively.

$$P_G := \sum_{k=\hat{k}}^{n} \binom{n}{k} r_G^k \left(1 - r_G\right)^{n-k}, \quad P_I := \sum_{k=\hat{k}}^{n} \binom{n}{k} r_I^k \left(1 - r_I\right)^{n-k}. \tag{6}$$

For each level of belief π, we denote the pair of equilibrium conviction probabilities of guilty defendants or innocent defendants by $\{(P_G, P_I)|\pi\}$. We also define correspondences of the conviction probabilities by

$$f_G(\pi) := \{P'_G | \, \exists P'_I, \, (P'_G, P'_I) \in \{(P_G, P_I) | \, \pi\}\},$$
$$f_I(\pi) := \{P'_I | \, \exists P'_G, \, (P'_G, P'_I) \in \{(P_G, P_I) | \, \pi\}\}.$$

Note that $f_G(.)$ and $f_I(.)$ are almost always single-valued. As we discussed after Proposition 1, efficient equilibrium voting behavior is almost always unique. The only exception is when $\pi = \bar{\pi}(\hat{k})$ under the unanimity rule.



(a) A super-majority rule ($\hat{k} = 8$). (b) The unanimity rule ($\hat{k} = 12$).

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

Proposition 2 (Properties of the Efficient Equilibrium Voting Behavior) *1. Convicting the guilty is no less likely than convicting the innocent: $P_G \geq P_I$ for all π.*

2. Efficient equilibrium voting behavior (σ_g, σ_i) is non-decreasing in π.

3. Conviction probabilities are non-decreasing in π : for every π and π' such that $\pi < \pi'$, $f_G(\pi) \leq f_G(\pi')$ and $f_I(\pi) \leq f_I(\pi')$.[8]

The above properties are derived from voting criteria (4) and (5). First, a juror with a guilty signal is more likely to vote for conviction ($\sigma_g \geq \sigma_i$). Since a guilty defendant is more likely to send guilty signals, jurors are more likely to vote for conviction ($r_G \geq r_I$) and a guilty defendant has a higher chance of being convicted ($P_G \geq P_I$). Second, given other jurors' voting behavior r_G and r_I, the value of the left-hand sides of both criteria are increasing in belief π. Thus, jurors with a higher π have higher incentives to vote for conviction. Last, the conviction probabilities are strictly increasing functions of σ_g and σ_i, which are in turn increasing correspondences of π. Therefore, the conviction probabilities $f_G(\cdot)$ and $f_I(\cdot)$ are increasing correspondences of π.

Figure 3 depicts the conviction probabilities induced by the efficient equilibrium voting behavior. The solid lines show the conviction probabilities of a guilty defendant; the dashed lines show the conviction probabilities of an innocent defendant.

[8]For sets A and B in $\mathbb{R}$, we denote by $A \leq B$ if $a \leq b$ for every $a \in A$ and $b \in B$.

4 Plea Bargaining

A prosecutor offers a defendant an opportunity to plead guilty and undergo the corresponding penalty $\theta \in [0, 1]$. The defendant compares θ with the conviction probability at trial induced by the efficient equilibrium voting behavior. A guilty defendant pleads guilty if $\theta \leq P_G$, and an innocent defendant pleads guilty if $\theta \leq P_I$.[9]

Recall that π denotes the jurors' common belief of the defendant's type conditional on a case proceeding to trial. When a case reaches a jury trial with a positive probability ($\phi_G < 1$ or $\phi_I < 1$), jurors update their common belief π by

$$\pi = \frac{\pi_0(1-\phi_G)}{\pi_0(1-\phi_G) + (1-\pi_0)(1-\phi_I)}. \tag{7}$$

If a defendant always pleads guilty ($\phi_G = \phi_I = 1$), we assume that the jurors have belief $\pi = 0$.[10]

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 trial. The pleading decisions lead jurors to update their belief about the defendants pool (π). This belief is taken as part of the evidence of guilt in the jury's voting behavior $\{(P_G, P_I)|\pi\}$. Conversely, the defendant takes into account the conviction probabilities in his pleading decision. Equilibrium behavior ensures that these interactions must be consistent with each other.

The following proposition summarizes the equilibrium restrictions.

Proposition 3 (Pleading Decisions and Voting Behavior) *Suppose the jury follows the efficient equilibrium voting behavior. For each prosecutor's offer θ, one, and only one, of the two following classes holds.*

1. ***Some guilty pleas**: Guilty defendants are indifferent between pleading guilty and undergoing a jury trial ($\theta = P_G$); innocent defendants prefer to plead not guilty ($\theta \geq$*

[9] We assume that defendants correctly anticipate conviction probabilities. In practice, participants in plea bargaining often foresee the outcomes of jury trials, so previous trial outcomes significantly influence the participants' bargaining power (see, e.g., Bibas (2004) and Stuntz (2004)). A novice defendant also gets advice from experienced defense attorneys.

[10] This refinement on jurors' common belief is equivalent to applying D1 by Cho and Kreps (1987) to the signaling game equilibrium induced by assuming that jurors follow the efficiency equilibrium voting behavior. Since a guilty defendant is more likely to be convicted at trial, if a defendant deviates from $\phi_G = \phi_I = 1$, he is more likely to be innocent. In such a case, D1 refines jurors' common belief π to be equal to 0.

P_I). $\theta = P_G \in f_G(\pi)$ for every equilibrium belief π.[11]

2. ***No guilty plea**: P_G, and necessarily P_I, are no more than θ. All defendants plead not guilty ($\phi_G = \phi_I = 0$). Thus, $\pi = \pi_0$ and $P_G \in f_G(\pi_0)$.*

In most cases, guilty defendants are indifferent between pleading guilty and pleading not guilty ($\theta = P_G$); innocent defendants prefer going to trial ($P_I \leq \theta$). To see why this occurs, suppose we have $\theta < P_G$. Guilty defendants will plead guilty, and depending on θ and P_I, only innocent defendants may go to trial. Then jurors believe all defendants in trials are innocent, and they will vote for acquittal ($P_G = 0$), contradicting $\theta < P_G$. Similarly, $\theta > P_G$ can hold in an equilibrium only when the prosecutor offers a very high level of θ. Then all defendants will go to trial, but the induced conviction probabilities (P_G and P_I) are still lower than θ.

The prosecutor wants to offer θ, which maximizes his expected equilibrium payoff, given the equilibrium restriction from Proposition 3. The prosecutor's problem is summarized by the following optimization problem:

$$\max_{\theta\in[0,1]} -\pi_0(1-q')\Big(\phi_G(1-\theta)+(1-\phi_G)(1-P_G)\Big)-(1-\pi_0)q'\Big(\phi_I\theta+(1-\phi_I)P_I\Big) \tag{8}$$

$$\text{subject to} \quad \begin{array}{ll} (a.1) & \phi_G \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_G, \\ (a.2) & \phi_I \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_I, \\ (b) & \pi = \begin{cases} 0 & \text{if } \phi_G=\phi_I=1 \\ \frac{\pi_0(1-\phi_G)}{\pi_0(1-\phi_G)+(1-\pi_0)(1-\phi_I)} & \text{otherwise,} \end{cases} \\ (c) & (P_G,P_I) \in \{(P'_G,P'_I)|\pi\}. \end{array}$$

The prosecutor's utility is decreased by $(1-q')$ when a guilty defendant goes without full punishment. Such a case is either the result of a guilty plea with probability ϕ_G and undelivered punishment $(1-\theta)$, or of acquittal in a jury trial with probability $(1-\phi_G)(1-P_G)$ and undelivered punishment 1. If an innocent defendant is mistakenly punished, the prosecutor's utility decreases with q'. The mistake is either the result of a guilty plea with probability ϕ_I and punishment θ, or of conviction in a jury trial with probability $(1-\phi_I)\,P_I$ and punishment 1. The equilibrium behavior of the defendant and the jurors restricts the

[11] The equilibrium belief π may not be unique. For instance, suppose that θ is equal to the conviction probability of a guilty defendant under $(\sigma_g = 1, \sigma_i = 0)$. Any π inducing $\sigma_g = 1$ and $\sigma_i = 0$ as the equilibrium voting behavior can be an equilibrium belief. Then, $f_G(\pi)$ contains $\theta = P_G$ and leads to the same level of equilibrium punishment.

prosecutor's optimization: a guilty or innocent defendant minimizes his expected punishment ($a.1$ and $a.2$), jurors rationally update their belief π (b), and jurors will follow the efficient equilibrium voting behavior (c).

The following proposition presents the prosecutor's optimal behavior and the consequent jurors' voting behavior. In the proposition, *some guilty pleas* and *no guilty plea* refer to the two classes of equilibrium behavior in Proposition 3.

Proposition 4 (Equilibrium Outcomes of Plea Bargaining and the Jury Trial) *1. If $q' > q$, the prosecutor induces **some guilty pleas**. Induced jury behavior resembles the behavior in the jury model without plea bargaining. However, jurors act as if they have the prosecutor's preferences q'.*

*2. If $q' \leq q$, the prosecutor induces **no guilty plea**. The jury behavior is the same as the behavior in the jury model without plea bargaining.*

To see the intuition behind Proposition 3, suppose the prosecutor sets θ such that $0 < \theta \leq \sup f_G(\pi_0)$, and thus $\theta = P_G > 0$.[12] Since the conviction probability of a guilty defendant is strictly between 0 and 1, the equilibrium voting behavior is responsive. Thus, we have $P_G > P_I$, which implies that an innocent defendant always goes to trial ($\phi_I = 0$). Using these equilibrium restrictions, we can simplify the prosecutor's objective function as

$$-\pi_0(1-q')(1-P_G) - (1-\pi_0)q'P_I. \tag{9}$$

We revisit the jurors' voting criteria and see how the prosecutor should influence the jurors' belief π. We modify (4) and (5) and obtain

$$\frac{Pr[\,piv\,|G]}{Pr[\,piv\,|I]}\frac{p}{1-p}\frac{\pi_0}{1-\pi_0} \quad \geq (\text{or } \leq) \quad \frac{q}{1-q}\frac{1-\pi}{\pi}\frac{\pi_0}{1-\pi_0} \quad \text{if the signal is } g,$$

and

$$\frac{Pr[\,piv\,|G]}{Pr[\,piv\,|I]}\frac{1-p}{p}\frac{\pi_0}{1-\pi_0} \quad \geq (\text{or } \leq) \quad \frac{q}{1-q}\frac{1-\pi}{\pi}\frac{\pi_0}{1-\pi_0} \quad \text{if the signal is } i.$$

That is, the jury behavior with belief π and the ratio of utility loses $\frac{q}{1-q}$ is equal to the jury behavior with belief π_0 and the ratio of utility loses equal to $\frac{q}{1-q}\frac{1-\pi}{\pi}\frac{\pi_0}{1-\pi_0}$. Thus, the prosecutor's effort to influence the jurors' belief has the same effect on jury behavior as

[12] The condition $\theta \leq \sup f_G(\pi_0)$ is necessary for $\theta = P_G$. If the prosecutor sets θ very high, all defendants proceed to jury trials, but P_G may still remain lower than θ.

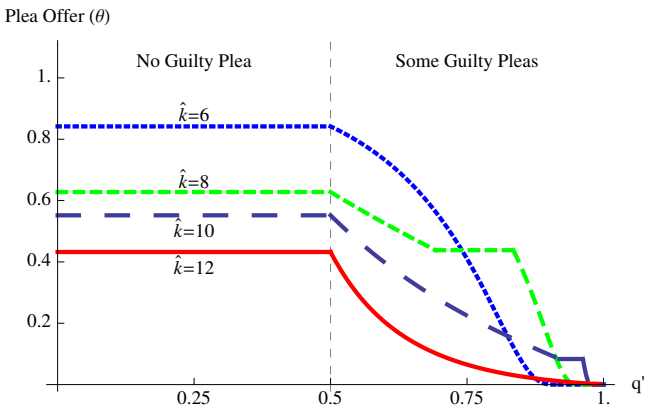


Figure 4: Optimal offer of guilty plea punishment: $n = 12$, $p = \frac{6}{10}$, $q = \frac{1}{2}$, and $\pi_0 = \frac{1}{2}$.

changing the level of the jurors' reasonable doubt, while keeping the jurors' belief equal to the prosecutor's initial belief π_0.

The prosecutor wants the jurors' induced ratio of utility loses $(\frac{q}{1-q}\frac{1-\pi}{\pi}\frac{\pi_0}{1-\pi_0})$ to perfectly coincide with the prosecutor's ratio of utility loses $(\frac{q'}{1-q'})$. Nevertheless, plea bargaining induces $\pi \leq \pi_0$, so the prosecutor can bias the jurors' reasonable doubt only against voting for conviction. If the jurors, rather than the prosecutor, already care more about punishing innocent defendants $(q > q')$, the prosecutor has no incentive to use plea bargaining for the purpose of influencing jury behavior.

Figure 4 illustrates the prosecutor's optimal offer of a guilty plea punishment. As Proposition 4 states, the optimal offer is divided into two classes. When the prosecutor is less concerned than jurors about punishing innocent defendants $(q' \leq q)$, the prosecutor offers a high level of punishment and induces a *no guilty plea*. If otherwise, the prosecutor offers a lower level of punishment and induces *some guilty pleas*. The more lenient punishment for a guilty plea leads to a higher proportion of guilty defendants pleading guilty. Then, jurors have a lower level of belief π, which eventually results in a lower chance of convicting an innocent defendant.

5 Comparison of Voting Rules

Our unified model of plea bargaining and jury trial can extend the scope of strategic jury models to the entire court process. Previously, applications of jury models were limited to jury trials only, which occurred in fewer than 5% of criminal cases in the United States. By

incorporating various jury models into our unified model, we can extend the insights from these models to the entire court process, including plea bargaining. We demonstrate this concept by considering Feddersen and Pesendorfer (1998).

Feddersen and Pesendorfer demonstrate the inferiority of the unanimity rule. As the number of jurors increases, the chance of convicting an innocent defendant and the chance of acquitting a guilty defendant do not converge to zero, whereas both chances converge to zero under any general super-majority rule. A reasonably sized jury would make judicial mistakes with a significant chance under the unanimity rule, but the chances are minimal under any super-majority rule.

The above observation holds in our unified model. Using the unanimity rule in a jury trial leads to inferior outcomes (i.e., punishment by guilty pleas and conviction probabilities) of the entire court process to those from any general super-majority rule.

Corollary 5 *(Comparing voting rules)*

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

2. *If the jury uses a general super-majority rule, the expected punishment for guilty defendants converges to one; the expected punishment for innocent defendants converges to zero.*

The inferiority of the unanimity rule remains after the addition of plea bargaining. This is because expected punishments are ultimately determined by conviction probabilities in a jury trial (Proposition 3), and the jury's behavior in our model is similar to the behavior in the jury model without plea bargaining (Proposition 4).

6 Conclusion

We highlight that the criminal court process implements the preferences of the party – prosecutor or jury – who is most concerned about wrongful conviction. The prosecutor chooses terms of a guilty plea, and the defendant in turn chooses whether to go to jury trial. Thus, if the prosecutor is more concerned about wrongful conviction than jurors, he can influence jury behavior through jurors' common belief about a defendant's guilt or

innocence at trials. Given our unified model of plea bargaining and jury trial, we also show that insights from previous jury models can be applied to the entire court process beyond trials.

Adding trial costs does not change the qualitative nature of the main results. Suppose the prosecutor can bring only a fraction of cases to trial, and defendants have heterogeneous trial costs. If defendants have relatively high trial costs, most innocent defendants, as well as most guilty defendants, will plead guilty. The large proportion of innocent defendants who take plea bargain offers is a concern of the prosecutor. But it is not a concern of the jurors because a jury makes a decision only for a single trial. Even if the prosecutor is less concerned about false conviction than jurors, the concern about mistakenly delivered punishment through plea bargaining may become so large that the prosecutor will propose play bargain offers more lenient than necessary to save trial costs. In this case, jurors are induced to vote more for acquittal.

Appendix

A Proof of Proposition 1

We first find all symmetric equilibrium voting behaviors (Appendix A.1), and then take the most efficient symmetric voting behavior (Appendix A.2).

A.1 Finding All Symmetric Equilibrium Voting Behaviors.

A.1.1 Non-responsive equilibrium voting behavior

We first consider the unanimity rule ($\hat{k} = n$). When all jurors always vote for conviction ($\sigma_g = 1, \sigma_i = 1$), being pivotal gives a single juror no information about other jurors private signals. Thus, the juror makes a voting decision only based on her own signal. When her signal is i, she votes for conviction (or acquittal) when

$$\frac{1-p}{p}\frac{\pi}{1-\pi} \geq \text{ (or } \leq) \frac{q}{1-q},$$

or equivalently

$$\pi \geq \text{ (or } \leq) \bar{\pi}(1).$$

In particular, always voting for conviction is an equilibrium behavior when $\pi \geq \bar{\pi}(1)$.

Now, consider general super-majority rules ($1 < \hat{k} < n$). Always voting for conviction ($\sigma_g = 1, \sigma_i = 1$) is an equilibrium voting behavior. When all jurors always vote for conviction, a single juror is never pivotal. Thus, no juror has an incentive to deviate from ($\sigma_g = 1, \sigma_i = 1$).

Similarly, always voting for acquittal ($\sigma_g = 0, \sigma_i = 0$) is an equilibrium voting behavior under both general super-majority rules and the unanimity rule.

A.1.2 Responsive equilibrium voting behavior

In a responsive equilibrium voting behavior ($\sigma_i < \sigma_g$), it must be that $0 < \sigma_g$ and $\sigma_i < 1$. So, the jury does not always deliver a guilty (or an innocent) verdict ($0 < r_G, r_I < 1$), and voting criteria (4) and (5) are well defined. There are three cases of responsive voting behavior: ($0 < \sigma_g < 1$, $\sigma_i = 0$), ($\sigma_g = 1$, $\sigma_i = 0$), and ($\sigma_g = 1$, $0 < \sigma_i < 1$). For each of these cases, we identify a range of consistent beliefs π.

Case 1: ($0 < \sigma_g < 1$, $\sigma_i = 0$) A juror with signal g is indifferent between voting for conviction and voting for acquittal:

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

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

$$\left(\frac{1-p\sigma_g}{1-(1-p)\sigma_g}\right)^{n-\hat{k}} \left(\frac{p}{1-p}\right)^{\hat{k}} \frac{\pi}{1-\pi} = \frac{q}{1-q}. \tag{10}$$

Under the unanimity rule ($\hat{k} = n$), the first term on the left-hand side is equal to 1. The equality holds when $\pi = \bar{\pi}(n)$, in which case any ($0 < \sigma_g < 1, \sigma_i = 0$) is an equilibrium voting behavior.

When the Equation (10) holds, $\sigma_g = 0$ implies $\pi = \bar{\pi}(\hat{k})$, and $\sigma_g = 1$ implies $\pi = \bar{\pi}(2\hat{k} - n)$. Moreover, $\frac{1-p\sigma_g}{1-(1-p)\sigma_g}$ is strictly decreasing in σ_g under any general super-majority rule ($1 < \hat{k} < n$). Thus, if there exists a responsive equilibrium behavior with $0 < \sigma_g < 1$, the belief π must be

$$\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(2\hat{k} - n).$$

For each π satisfying the above condition, we find a unique equilibrium voting behavior ($\sigma_g, \sigma_i = 0$) through algebraic manipulation of (10):

$$\sigma_g(\pi) = \frac{\psi_1 - 1}{(1-p)\psi_1 - p} \quad \text{where} \quad \psi_1 = \left(\frac{1-p}{p}\right)^{\frac{\hat{k}}{n-\hat{k}}} \left(\frac{q}{1-q}\,\frac{1-\pi}{\pi}\right)^{\frac{1}{n-\hat{k}}}. \tag{11}$$

Case 2: $(\sigma_g = 1,\ \sigma_i = 0)$ A juror with signal g prefers to vote for conviction, whereas a juror with signal i prefers to vote for acquittal. We substitute in $r_G = p$ and $r_I = 1 - p$ to voting criteria (4) and (5) and obtain

$$\left(\frac{p}{1-p}\right)^{2(\hat{k}-1)-n} \leq \frac{q}{1-q}\,\frac{1-\pi}{\pi} \leq \left(\frac{p}{1-p}\right)^{2\hat{k}-n}. \tag{12}$$

The above inequalities are equivalent to

$$\bar{\pi}(2\hat{k} - n) \leq \pi \leq \bar{\pi}(2(\hat{k} - 1) - n).$$

That is, $(\sigma_g = 1, \sigma_i = 0)$ is an equilibrium voting behavior when π is between $\bar{\pi}(2\hat{k} - n)$ and $\bar{\pi}(2(\hat{k} - 1) - n)$.

Case 3: $(\sigma_g = 1,\ 0 < \sigma_i < 1)$ A juror with signal i is indifferent between voting for conviction and voting for acquittal:

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

We substitute in $r_G = p + (1-p)\sigma_i$ and $r_I = (1-p) + p\sigma_i$ and obtain

$$\left(\frac{p + (1-p)\sigma_i}{(1-p) + p\sigma_i}\right)^{\hat{k}-1} \left(\frac{1-p}{p}\right)^{n-\hat{k}+1} \frac{1-\pi}{\pi} = \frac{q}{1-q}. \tag{13}$$

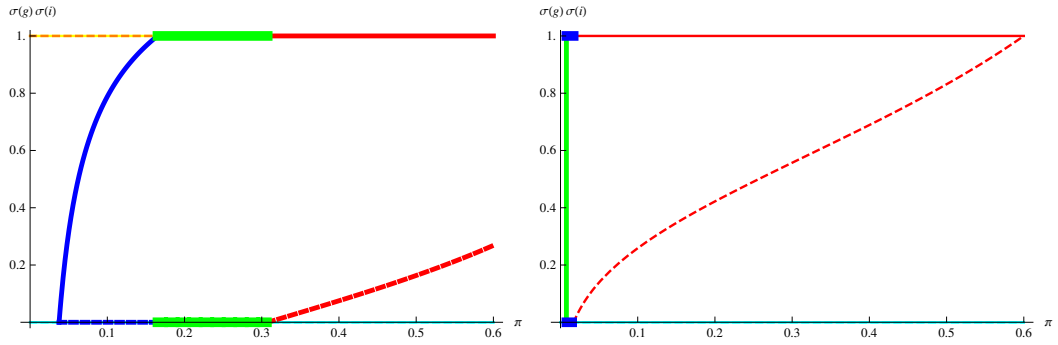
When the Equation (13) holds, $\sigma_i = 0$ implies $\bar{\pi}(2(\hat{k} - 1) - n) = \pi$, and $\sigma_i = 1$ implies $\pi = \bar{\pi}(-n + \hat{k} - 1)$. Moreover, $\frac{p+(1-p)\sigma_i}{(1-p)+p\sigma_i}$ is strictly decreasing in σ_i since $\hat{k} > 1$. Thus, if there exists a responsive equilibrium voting behavior with $\sigma_g = 1$ and $0 < \sigma_i < 1$, π must satisfy

$$\bar{\pi}(2(\hat{k} - 1) - n) < \pi < \bar{\pi}(-n + \hat{k} - 1).$$

For each π satisfying the above inequalities, we find a unique equilibrium voting behavior

General super-majority rules ($1 < \hat{k} < n$)		The unanimity rule ($\hat{k} = n$)	
Non-responsive voting			
$\forall\ \pi \in [0, \pi_0]$	$(\sigma_g = \sigma_i = 1)$	$\pi \geq \pi(1)$	$(\sigma_g = \sigma_i = 1)$
$\pi \in [0, \pi_0]$	$(\sigma_g = \sigma_i = 0)$	$\forall\ \pi \in [0, \pi_0]$	$(\sigma_g = \sigma_i = 0)$
Responsive voting			
$\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(2\hat{k} - n)$	$(0 < \sigma_g < 1, \sigma_i = 0)$	$\pi = \bar{\pi}(n)$	$(0 < \sigma_g < 1, \sigma_i = 0)$
$\bar{\pi}(2\hat{k} - n) \leq \pi \leq \bar{\pi}(2(\hat{k} - 1) - n)$	$(\sigma_g = 1, \sigma_i = 0)$	$\bar{\pi}(n) \leq \pi \leq \bar{\pi}(n-2)$	$(\sigma_g = 1, \sigma_i = 0)$
$\bar{\pi}(2(\hat{k} - 1) - n) < \pi < \bar{\pi}(-n + \hat{k} - 1)$	$(\sigma_g = 1, 0 < \sigma_i < 1)$	$\bar{\pi}(2n-2) < \pi < \bar{\pi}(-1)$	$(\sigma_g = 1, 0 < \sigma_i < 1)$

Table 1: Symmetric voting equilibrium behavior in jury trial.



(a) A super-majority rule ($\hat{k} = 8$). (b) The unanimity rule ($\hat{k} = 12$).

Figure 5: Symmetric equilibrium voting behavior with $n = 12$, $p = \frac{6}{10}$, and $q = \frac{6}{10}$

$(\sigma_g = 1, \sigma_i)$ through algebraic manipulation of (13):

$$\sigma_i(\pi) = \frac{p - \psi_2(1-p)}{p\,\psi_2 - (1-p)} \quad \text{where} \quad \psi_2 = \left(\frac{p}{1-p}\right)^{\frac{n-\hat{k}+1}{\hat{k}-1}} \left(\frac{q}{1-q}\,\frac{1-\pi}{\pi}\right)^{\frac{1}{\hat{k}-1}}. \tag{14}$$

Table 1 summarizes all equilibrium voting behaviors. Figure 5 illustrates the equilibria with $n = 12$, $p = \frac{6}{10}$, and $q = \frac{6}{10}$ under voting rules $\hat{k} = 8$ and $\hat{k} = 12$. We use solid lines for σ_g and dashed lines for σ_i. The pair of σ_g and σ_i sharing the same thickness in Figure 5 forms a strategy profile.

A.2 Finding an Efficient Equilibrium Voting Behavior

When there is no responsive equilibrium voting behavior ($\pi \leq \bar{\pi}(\hat{k})$), one of the non-responsive equilibria is an efficient equilibrium voting behavior.

Note that jurors' utility is

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

Thus, the expected utility from $(\sigma_g = \sigma_i = 0)$ is higher than the utility from $(\sigma_g = \sigma_i = 1)$ if and only if $q(1-\pi) \geq (1-q)\pi$: i.e., $\pi \leq q$.

If responsive equilibrium voting behavior exists, it is more efficient than both non-responsive voting behaviors. We illustrate this result by showing that the payoff from responsive voting is higher than the payoff from $(\sigma_g = 0, \sigma_i = 0)$. We omit the comparison between the payoff from responsive voting and the payoff from $(\sigma_g = 1, \sigma_i = 1)$.

Note that responsive voting is more efficient than $(\sigma_g = 0, \sigma_i = 0)$ if and only if the conviction probabilities (P_G, P_I) of responsive voting behavior satisfy

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

We rewrite the inequality and obtain

$$\frac{P_G}{P_I} = \frac{\sum_{j=\hat{k}}^{n} \binom{n}{j} r_G^j (1-r_G)^{n-j}}{\sum_{j=\hat{k}}^{n} \binom{n}{j} r_I^j (1-r_I)^{n-j}} > \text{ (or } \geq) \frac{q}{1-q}\frac{1-\pi}{\pi}. \tag{15}$$

When the above inequality holds as a strict inequality, the responsive equilibrium voting behavior is strictly more efficient than $(\sigma_g = 0, \sigma_i = 0)$. If the inequality holds as an equality, the two equilibrium voting behaviors are equally efficient.

We proceed separately with general super-majority rules and the unanimity rule.

A.2.1 General super-majority rules ($1 < \hat{k} < n$)

First, $k' > k$ and $r_G > r_I > 0$ imply

$$\frac{r_G^{k'}(1-r_G)^{n-k'}}{r_I^{k'}(1-r_I)^{n-k'}} > \frac{r_G^{k}(1-r_G)^{n-k}}{r_I^{k}(1-r_I)^{n-k}}. \tag{16}$$

And $x, x' > 0$ and $y, y' > 0$ imply

$$\frac{x'}{y'} > \frac{x}{y} \quad \text{implies} \quad \frac{x+x'}{y+y'} > \frac{x}{y}. \tag{17}$$

Sequentially applying (16) for $(\hat{k}+1, \hat{k}), \ldots, (n, \hat{k})$, and using (17), we obtain

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

To prove that (15) holds as a strict inequality, it is enough to show that

$$\frac{r_G^{\hat{k}}(1-r_G)^{n-\hat{k}}}{r_I^{\hat{k}}(1-r_I)^{n-\hat{k}}} \geq \frac{q}{1-q}\,\frac{1-\pi}{\pi}. \tag{18}$$

We proceed with each of three cases of responsive equilibrium voting behavior.

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

Since $r_G = p\sigma_g$ and $r_I = (1-p)\sigma_g$, the left-hand side of (18) is

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

The equilibrium restriction (10) implies that the right-hand side of the above equation is equal to the right-hand side of (18). Thus (18) holds as an 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 left-hand side of (18) is

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

Equation (12) implies that Equation (18) must be true.

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

From (13), we have

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

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

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

The above equality and inequality imply (18).

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

If the voting rule follows the unanimity rule, (15) becomes

$$\frac{P_G}{P_I} = \left(\frac{r_G}{r_I}\right)^n > \text{ (or } \geq) \frac{q}{1-q}\frac{1-\pi}{\pi}. \tag{19}$$

We again proceed to show (19) for each of three cases of responsive equilibrium voting behavior.

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 left-hand side of (19) becomes

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

Since $\pi = \bar{\pi}(n)$, (19) holds as an equality. Therefore, both the responsive equilibrium voting behavior 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 left-hand side of (19) becomes

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

When $\pi = \bar{\pi}(2\hat{k} - n) = \bar{\pi}(n)$, (19) holds as an equality. If $\bar{\pi}(n) < \pi \leq \bar{\pi}(2(\hat{k} - 1) - n)$, (19) holds as 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 more efficient than non-responsive voting $(\sigma_g = \sigma_i = 0)$.

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

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

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

The last equation is from the voting criterion (13). Responsive voting is more efficient than non-responsive voting $(\sigma_g = \sigma_i = 0)$.

B Proof of Proposition 2.

Let (σ_g, σ_i) be the efficient equilibrium voting behavior, and $r_G = p\sigma_g + (1-p)\sigma_i$ and $r_I = (1-p)\sigma_g + p\sigma_i$. The conviction probabilities are determined by

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

Item 1: When the efficient equilibrium voting behavior is non-responsive, $P_G \geq P_I$ clearly holds. If the efficient equilibrium voting behavior is responsive, $P_G > P_I$ holds since a guilty defendant is more likely to send a guilty signal, and a juror is more likely to vote for conviction with a guilty signal ($\sigma_g > \sigma_i$).

Item 2: We observed that σ_g and σ_i are non-decreasing in π over $(\bar{\pi}(\hat{k}), \bar{\pi}(2\hat{k}-n))$ and $(\bar{\pi}(2\hat{k}-n), \bar{\pi}(-n+\hat{k}-1)]$, and otherwise constant in π. Thus, σ_g and σ_i are non-decreasing in π over $[0, \pi_0]$.

Item 3: The conviction probabilities are strictly increasing in σ_g and σ_i, and σ_g and σ_i are non-decreasing in π. Thus, $f_G(\pi)$ and $f_I(\pi)$ are non-decreasing in π.

C Proof of Proposition 3

We first show that $f_G(\pi)$ is upper hemicontinuous in π with non-empty convex values. f_G is continuous in σ_g and σ_i and the efficient equilibrium behaviors (σ_g and σ_i) are upper hemicontinuous in π. Therefore, $f_G(\pi)$ is upper hemicontinuous in π. In addition, $f_G(\pi)$ is convex-valued for all π. The efficient equilibrium behavior $\sigma_g, \sigma_i)$ is unique for almost every π, and thus $f_G(\pi)$ is convex-valued. The only exception is when $\pi = \bar{\pi}(n)$ and the rule is unanimous. In that case, the efficient equilibrium voting behavior is any pair (σ_g, σ_i) satisfying $\sigma_i = 0$ and $0 \leq \sigma_g \leq 1$. $\sum_{k'=\hat{k}}^{n} \binom{n}{k'} r_G^{k'} (1-r_G)^{n-k'}$ is continuous in (σ_g, σ_i), so $f_G(\pi)$ is convex-valued even if the efficient equilibrium behavior is not unique.

To prove the first item in Proposition 3, suppose $\theta \leq P_G$. It is necessary that $\theta \in [0, \bar{\theta}]$ where $\bar{\theta} = \sup f_G(\pi_0)$. If $\theta < P_G$, every guilty defendant pleads guilty, and only innocent defendants may or may not go to trial. In such a case, jurors believe that all defendants in trials are innocent ($\pi = 0$), which leads the conviction probability to equal zero. This contradicts $\theta < P_G$, and it must be that $\theta = P_G$. Since $f_G(\pi)$ is upper hemicontinuous in π with non-empty convex values, there exists a π such that $\theta \in f_G(\pi)$ (Intermediate Value

Theorem).

If we instead have $\theta > P_G$, no defendant pleads guilty, and the jurors believe $\pi = \pi_0$. The conviction probabilities (P_G, P_I) must be in $\{(P'_G, P'_I)|\pi_0\}$, which is the case stated in the second item in Proposition 3.

D Proof of Proposition 4

D.1 Simplifying the Prosecutor's Problem

The prosecutor's problem is

$$\max_{\theta\in[0,1]} -\pi_0(1-q')\Big(\phi_G(1-\theta)+(1-\phi_G)(1-P_G)\Big)-(1-\pi_0)q'\Big(\phi_I\theta+(1-\phi_I)P_I\Big) \tag{20}$$

$$\text{subject to} \quad \begin{array}{ll} (a.1) & \phi_G \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_G \\ (a.2) & \phi_I \in \arg\min_{\phi'\in[0,1]} \phi'\theta+(1-\phi')P_I \\ (b) & \pi = \begin{cases} 0 & \text{if } \phi_G=\phi_I=1 \\ \frac{\pi_0(1-\phi_G)}{\pi_0(1-\phi_G)+(1-\pi_0)(1-\phi_I)} & \text{otherwise.} \end{cases} \\ (c) & (P_G,P_I)\in\{(P'_C,P'_I)|\pi\}. \end{array}$$

We simplify the prosecutor's problem using Proposition 3.

It is without loss of generality to assume that $\theta \in [0, \bar{\theta}]$ because the prosecutor can obtain any utility from offering $\theta > \bar{\theta}$ by offering $\theta = \bar{\theta}$. If $\theta > \bar{\theta}$, a defendant always pleads not guilty and anticipates conviction probabilities $(P_G, P_I) \in \{(P'_G, P'_I)|\pi_0\}$. As the prosecutor sets $\theta = \bar{\theta}$, the prosecutor and a guilty defendant are indifferent between pleading guilty and pleading not guilty. The punishment for a guilty plea is equal to the conviction probability, which is equal to the expected value of punishment from going to a jury trial.

If the prosecutor offers $\theta \in [0, \bar{\theta}]$, the prosecutor can assume that $\phi_I = 1$. When $\theta \in [0, \bar{\theta}]$, we have $\theta = P_G \geq P_I$ (Proposition 3). In general, we have $\theta = P_G > P_I$, so an innocent defendant will not plead guilty ($\phi_I = 1$). If $\theta = P_G = P_I = 0$, any pleading decision, including the case $\phi_I = 1$, incurs the same expected utility $-\pi_0(1 - q')$ for the prosecutor.

By applying the above observations, we simplify the prosecutor's problem as

$$\max_{\theta\in[0,\bar{\theta}]} -\pi_0(1-q')(1-\theta)-(1-\pi_0)q'P_I$$

$$\text{subject to} \quad \begin{array}{ll} (a) & \phi_G \in [0,1] \\ (b) & \pi = \begin{cases} 0 & \text{if } \phi_G = 1 \\ \frac{\pi_0(1-\phi_G)}{\pi_0(1-\phi_G)+(1-\pi_0)} & \text{otherwise.} \end{cases} \\ (c) & (\theta, P_I) \in \{(P_G', P_I')|\pi\}. \end{array}$$

We simplify the prosecutor's problem even further.

Define $\tilde{P}_I : [0,\bar{\theta}] \to [0,1]$ as follows:

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

The function $\tilde{P}_I$ is well-defined: the value of $\tilde{P}_I(\theta)$ exists and is unique for every $\theta \in [0,\bar{\theta}]$. Let $\hat{p_G}$ be the conviction probability of a guilty defendant when jurors vote by following their own signals ($\sigma_g = 1, \sigma_i = 0$). Consider four cases: (1) $\theta = 0$, (2) $\theta \in (0, \hat{p_G})$, (3) $\theta = \hat{p_G}$, and (4) $\theta \in (\hat{p_G}, \bar{\theta}]$. If $\theta = 0$, then $\tilde{P}_I(\theta)$ must be 0. If $\theta = \hat{p_G}$, then $\tilde{P}_I(\theta)$ has a unique value that is derived by the voting strategy ($\sigma_g = 1, \sigma_i = 0$).

For other cases, $\tilde{P}_I(\theta)$ is determined by

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

where $r_G = p\sigma_g + (1-p)\sigma_i$ and $r_I = (1-p)\sigma_g + p\sigma_i$.

When $\theta \in (0, \hat{p_G})$, we have $\sigma_i = 0$, and both P_G and P_I are strictly increasing in σ_g. For any $\theta \in (0, \hat{p_G})$, there exists a unique σ_g inducing $P_G = \theta$ because P_G is continuous in r_G, which is continuous in σ_g. The 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 $\theta \in (\hat{p_G}, \bar{\theta}]$.

We also verify that $\tilde{P}_I$ is strictly increasing and continuous in θ on $[0,\bar{\theta}]$, and differentiable on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$. By using $\tilde{P}_I(\theta)$, we simplify the prosecutor's problem as

$$\max_{\theta\in[0,\bar{\theta}]} U(\theta) := -\pi_0(1-q')(1-\theta)-(1-\pi_0)q'\tilde{P}_I(\theta). \tag{21}$$

In the next subsection, we show that the objective function $U(\theta)$ is strictly concave in θ. So, the first-order conditions are necessary and sufficient to characterize the maximizer θ^*.

We later investigate the first-order conditions to prove Proposition 4.

D.2 $U(\theta)$ is strictly concave in θ.

The objective function $U(\theta)$ is continuous in θ because $\tilde{P}_I$ is continuous. Moreover, $U(\theta)$ is differentiable with respect to θ on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$ because $\tilde{P}_I$ is differentiable on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$, and $U(\theta)$ is a linear combination of θ and $\tilde{P}_I$.

We lastly show that $U(\theta)$ is concave on $[0, \bar{\theta}]$. Since $U(\theta)$ is a linear combination of θ and $\tilde{P}_I$, it is enough to prove that $\tilde{P}_I$ is concave on $[0, \theta]$. In particular, we show that the derivative of $\tilde{P}_I$ is decreasing on $(0, \hat{p_G})$ and $(\hat{p_G}, \bar{\theta})$, and the left derivative at $\theta = \hat{p_G}$ is greater than the right derivative.

If $\theta \in (0, \hat{p_G})$, P_G and P_I are differentiable with respect to σ_g:

$$\begin{aligned}
\frac{\partial P_G}{\partial \sigma_g} &= \frac{\partial}{\partial \sigma_g} \sum_{k=\hat{k}}^{n} \binom{n}{k} (r_G)^k (1 - r_G)^{n-k} \\
&= \sum_{k=\hat{k}}^{n-1} \left(\frac{n!}{k!(n-k)!} k r_G^{k-1} (1 - r_G)^{n-k} r_G' \right. \\
&\quad \left. - \frac{n!}{k!(n-k-1)!} r_G^k (n-k)(1 - r_G)^{n-k-1} r_G' \right) + n r_G^{n-1} r_G' \\
&= n\, r_G' \binom{n-1}{\hat{k}-1} r_G^{\hat{k}-1} (1 - r_G)^{n-\hat{k}}. \qquad (22)
\end{aligned}$$

Similarly,

$$\frac{\partial P_I}{\partial \sigma_g} = n\, r_I' \binom{n-1}{\hat{k}-1} r_I^{\hat{k}-1} (1 - r_I)^{n-\hat{k}}. \qquad (23)$$

Therefore,

$$\frac{\partial \tilde{P}_I(\theta)}{\partial \theta} = \frac{\partial P_I / \partial \sigma_g}{\partial P_G / \partial \sigma_g} = \frac{r_I'\, r_I^{\hat{k}-1} (1 - r_I)^{n-\hat{k}}}{r_G'\, r_G^{\hat{k}-1} (1 - r_G)^{n-\hat{k}}}. \qquad (24)$$

Since $r_G = p\sigma_g$ and $r_I = (1-p)\sigma_g$, (24) becomes

$$\frac{\partial \tilde{P}_I(\theta)}{\partial \theta} = \left(\frac{1-p}{p} \right)^{\hat{k}} \left(\frac{1 - (1-p)\sigma_g}{1 - p\sigma_g} \right)^{n-\hat{k}}. \qquad (25)$$

Note that σ_g is increasing in $\theta \in (0, \hat{p_G})$. Thus, the above derivative is strictly decreasing in $\theta \in (0, \hat{p_G})$. That is, $\frac{\partial \tilde{P}_I(\theta)}{\partial \theta}$ is decreasing in $\theta \in (0, \hat{p_G})$, so $\tilde{P}_I(\theta)$ is strictly concave in $(0, \hat{p_G})$.

If $\theta \in (\hat{p_G}, \bar{\theta})$, $\sigma_g = 1$ and σ_i varies. Similar to (22) and (23), we obtain

$$\frac{\partial \tilde{P}_I(\theta)}{\partial \theta} = \frac{\partial P_I / \partial \sigma_i}{\partial P_G / \partial \sigma_i} = \frac{r_I' \, r_I^{\hat{k}-1}(1-r_I)^{n-\hat{k}}}{r_G' \, r_G^{\hat{k}-1}(1-r_G)^{n-\hat{k}}}. \tag{26}$$

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

$$\frac{\partial \tilde{P}_I(\theta)}{\partial \theta} = \left(\frac{(1-p) + p\sigma_i}{p + (1-p)\sigma_i}\right)^{\hat{k}-1} \left(\frac{p}{1-p}\right)^{n-\hat{k}+1}. \tag{27}$$

Note that σ_i is increasing in $\theta \in (\hat{p_G}, \bar{\theta})$. The above derivative is strictly decreasing in $\theta \in (\hat{p_G}, \bar{\theta})$. Thus, $\tilde{P}_I(\theta)$ is strictly concave in $(\hat{p_G}, \bar{\theta})$.

Last, the left derivative of $\tilde{P}_I$ is greater than the right derivative at $\theta = \hat{p_G}$ because the limit of (25) as σ_g goes to 1 is greater than the limit of (27) as σ_i goes to 0. This completes the proof that $\tilde{P}_I$ is strictly concave in $\theta \in [0, \bar{\theta}]$.

D.3 First-Order Conditions

The prosecutor's objective function is strictly concave in θ. Then first-order conditions are necessary and sufficient for characterizing the optimizer θ^*. We prove Proposition 4 by investigating first-order conditions.

D.3.1 Interior solutions

- $(0 < \theta^* < \hat{p_G})$: Note that the first-order condition $\frac{\partial U(\theta)}{\partial \theta} = 0$ at $\theta = \theta^*$ implies that

$$\left(\frac{p}{1-p}\right)^{\hat{k}} \left(\frac{1 - p\sigma_g}{1 - (1-p)\sigma_g}\right)^{n-\hat{k}} = \frac{q'}{1-q'} \frac{1-\pi_0}{\pi_0}.$$

  A juror with signal g uses a mixed strategy in the efficient equilibrium voting behavior induced by θ^*. So, Equation (10) holds, which implies

$$\frac{q}{1-q} \frac{1-\pi}{\pi} = \frac{q'}{1-q'} \frac{1-\pi_0}{\pi_0}.$$

- $(\hat{p_G} < \theta^* < \bar{\theta})$: The first-order condition (21) becomes

$$\left(\frac{p + (1-p)\sigma_i}{(1-p) + p\sigma_i}\right)^{\hat{k}-1} \left(\frac{1-p}{p}\right)^{n-\hat{k}+1} = \frac{q'}{1-q'} \frac{1-\pi_0}{\pi_0}.$$

A juror with signal i uses a mixed strategy in the efficient equilibrium voting behavior induced by θ^*. Thus, Equation (14) holds, which implies

$$\frac{q}{1-q}\frac{1-\pi}{\pi} = \frac{q'}{1-q'}\frac{1-\pi_0}{\pi_0}.$$

D.3.2 Boundary solutions

- $(\theta^* = \hat{p_G})$: Since $\hat{p_G}$ is a unique maximizer of (21), we have

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

By rewriting $\frac{\partial \tilde{P}_I(\theta)}{\partial \theta}$ using (25) and (27), we obtain

$$\left(\frac{(1-p)+p\sigma_i}{p+(1-p)\sigma_i}\right)^{\hat{k}-1}\left(\frac{p}{1-p}\right)^{n-\hat{k}+1} \leq \frac{1-q'}{q'}\frac{\pi_0}{1-\pi_0} \leq \left(\frac{1-p}{p}\right)^{\hat{k}}\left(\frac{1-(1-p)\sigma_g}{1-p\sigma_g}\right)^{n-\hat{k}},$$

which implies that

$$\left(\frac{p}{1-p}\right)^{2(\hat{k}-1)-n} \leq \frac{q'}{1-q'}\frac{1-\pi_0}{\pi_0} \leq \left(\frac{p}{1-p}\right)^{2\hat{k}-n}.$$

We compare the above inequalities with (12). The jurors' voting behavior with π and q is exactly the same as the voting behavior when the jurors' belief is π_0 and their reasonable doubt is q'.

- $(\theta^* = 0)$: The right derivative at $\theta = 0$ must be less than or equal to 0. We apply (25) to the derivative of the objective function in (21) and take σ_g close to 0. Then

$$\left(\frac{p}{1-p}\right)^{\hat{k}} \leq \frac{q'}{1-q'}\frac{1-\pi_0}{\pi_0}.$$

Note that θ^* induces the equilibrium voting behavior $(\sigma_g = 0, \sigma_i = 0)$. This strategy profile becomes an efficient equilibrium voting behavior when the right-hand side of (10) is greater than or equal to the left-hand side. That is,

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

From the above two inequalities, we observe that the equilibrium voting behavior is the same when jurors' belief is π_0 and their reasonable doubt is q'.

- $(\theta^* = \bar{\theta})$: The left derivative at $\theta = \bar{\theta}$ must be non-negative. By applying (27) to the derivative of $U(\theta)$, we obtain

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

which implies that

$$\left(\frac{p + (1-p)\bar{\sigma}_i}{(1-p) + p\bar{\sigma}_i}\right)^{\hat{k}-1} \left(\frac{1-p}{p}\right)^{n-\hat{k}+1} \geq \frac{q'}{1-q'} \frac{1-\pi_0}{\pi_0},$$

where $(\bar{\sigma}_i, \sigma_g = 1)$ is an equilibrium voting behavior at $\pi = \pi_0$.

In this situation, a juror with signal i is indifferent between voting for conviction and voting for acquittal. Thus, (13) becomes

$$\left(\frac{p + (1-p)\bar{\sigma}_i}{(1-p) + p\bar{\sigma}_i}\right)^{\hat{k}-1} \left(\frac{1-p}{p}\right)^{n-\hat{k}+1} = \frac{q}{1-q} \frac{1-\pi_0}{\pi_0}.$$

From the above two expressions, we have

$$\frac{q}{1-q} \geq \frac{q'}{1-q'},$$

which implies

$$q \geq q'.$$

If $q \geq q'$, the prosecutor offers $\theta^* = \bar{\theta}$, and all defendants plead not guilty ($\pi = \pi_0$). Jurors' reasonable doubt is $\frac{q}{1-q}$, which is the same as the reasonable doubt in the 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 equilibrium expected utility for the prosecutor as $\theta^* = \bar{\theta}$.

D.4 Proof of Corollary 5

Fix $\alpha \in (0, 1)$ and a jury size n, let $\hat{k}$ be the smallest integer that is larger than αn. An efficient equilibrium voting behavior is responsive for every $\bar{\pi}(\hat{k}) < \pi < \bar{\pi}(-n + \hat{k} - 1)$. Note that $\bar{\pi}(l)$ is by definition strictly decreasing in l. Thus, a responsive voting behavior becomes

the efficient equilibrium voting behavior for every $0 < \pi \leq 1$ the jury size n increases to infinity.

Consider the efficient equilibrium voting behavior under the unanimity rule. The conviction probability of a guilty defendant converges to $1 - \left(\frac{(1-q)(1-p)\pi}{qp(1-\pi)}\right)^{\frac{1-p}{2p-1}}$, and the conviction probability of an innocent defendant converges to $\left(\frac{(1-q)(1-p)\pi}{qp(1-\pi)}\right)^{\frac{p}{2p-1}}$. (See Proposition 2 in Feddersen and Pesendorfer (1998))

Under a general super-majority rule, for every jury size n, we have $\frac{1-\pi}{\pi} = \frac{1-\pi_0}{\pi_0}$ if $q > q'$, or $\frac{q}{1-q}\frac{1-\pi}{\pi} = \frac{q'}{1-q'}\frac{1-\pi_0}{\pi_0}$ if $q \leq q'$. Thus, we define

$$\tilde{q} = \max\{q, q'\},$$

and obtain

$$\frac{q}{1-q}\frac{1-\pi}{\pi} = \frac{\tilde{q}}{1-\tilde{q}}\frac{1-\pi_0}{\pi_0}.$$

We now apply Proposition 3 in Feddersen and Pesendorfer (1998): the conviction probability for a guilty defendant converges to 1 and for an innocent defendant to 0.

From Proposition 3 in this paper, we can relate the expected punishments, one for guilty defendants and the other for innocent defendants, to the conviction probabilities in jury trials.

E Other Notions of Equilibrium Refinements.

It is worth investigating if we can apply equilibrium refinement concepts other than the most efficient equilibrium. It turns out that refinement using *weakly undominated strategies* by Gerardi and Yariv (2007) or the *trembling hand perfection* by Austen-Smith and Feddersen (2005) does not lead to a unique equilibrium voting behavior at every level of jurors' common belief π.

Take any super-majority rule, and suppose that $1 < \hat{k} < n$ and $0 < \pi < \bar{\pi}(\hat{k})$. Previously in the proof of Proposition 1, we observed that $(\sigma_g = 1, \sigma_i = 1)$ and $(\sigma_g = 0, \sigma_i = 0)$ are symmetric equilibria. We demonstrate that both $(\sigma_g = 0, \sigma_i = 0)$ and $(\sigma_g = 1, \sigma_i = 1)$ are weakly undominated strategies and neither of them passes the trembling hand perfection criterion.

First, $(\sigma_g = 0, \sigma_i = 0)$ is not a weakly dominated strategy. Suppose that all other jurors, except juror j, play $(\sigma'_g = 1, \frac{1}{2} < \sigma'_i < 1)$. If juror j is pivotal, $\hat{k} - 1$ other jurors

vote for conviction. The best response for juror j with signal g is to vote for acquittal. Some other jurors' conviction votes may come from signal i. Thus, other $\hat{k}-1$ jurors' votes for conviction combined with juror j's guilty signal give insufficient evidence to overcome $\pi < \bar{\pi}(\hat{k})$. Clearly, the best response for a juror with signal i is also voting for acquittal. Therefore, $(\sigma_g = 0, \sigma_i = 0)$ is not a weakly dominated strategy.

Second, $(\sigma_g = 1, \sigma_i = 1)$ is not a weakly dominated strategy either. Suppose all other jurors, except juror j, play $(0 < \sigma_g'' < \frac{1}{2}, \sigma_i'' = 0)$. If juror j is pivotal, $\hat{k}-1$ other jurors vote for conviction. The best response for juror j with signal i is to vote for acquittal. Other $\hat{k}-1$ jurors' conviction votes may offer even stronger evidence that the defendant is guilty than the evidence from $\hat{k}$ guilty signals. The best response for juror j is to vote for conviction regardless of her own signal. Therefore, $(\sigma_g = 1, \sigma_i = 1)$ is not a weakly dominated strategy.

Neither $(\sigma_g = 0, \sigma_i = 0)$ nor $(\sigma_g = 1, \sigma_i = 1)$ passes the trembling hand perfection. Suppose that $(\sigma_g = 0, \sigma_i = 0)$ is a Bayesian Nash equilibrium satisfying the trembling hand perfection. That is, there exists a sequence of perturbed games such that a corresponding sequence of Bayesian Nash equilibria $(\sigma_g^n = \epsilon_1^n, \sigma_i^n = \epsilon_2^n)$ assigns strictly positive probabilities to both pure strategies and converges to $(\sigma_g = 0, \sigma_i = 0)$. However, such a sequence of Bayesian Nash equilibria can not exist. A guilty signal g gives a strictly higher incentive to vote for conviction than an innocent signal i. If a juror with signal g is indifferent between voting for conviction and voting for acquittal, she must strictly prefer to vote for acquittal with signal i. Therefore, $(\sigma_g = 0, \sigma_i = 0)$ does not pass the trembling hand perfection. Similarly, $(\sigma_g = 1, \sigma_i = 1)$ does not pass the trembling hand perfection criterion, either.

References

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

Austen-Smith, D., and T. Feddersen (2005): "Deliberation and Voting Rules," *Social Choice and Strategic Decisions*, pp. 269–316.

Bac, M. (2000): "Signaling bargaining power: Strategic delay versus restricted offers," *Economic Theory*, 16(1), 227–237.

Baker, S., and C. Mezzetti (2001): "Prosecutorial Resources, Plea Bargaining, and the Decision to Go to Trial," *Journal of Law, Economics, and Organization*, 17(1), 149–167.

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

Bjerk, D. (2007): "Guilt Shall not Escape or Innocence Suffer? The Limits of Plea Bargaining When Defendant Guilt is Uncertain," *American Law and Economics Review*, 9(2), 305–329.

Busch, L.-A., and I. J. Horstmann (1999): "Signaling via an agenda in multi-issue bargaining with incomplete information," *Economic Theory*, 13(3), 561–575.

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

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," *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," *American Political Science Review*, 92(1), 23–35.

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

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

Grossman, G., and M. Katz (1983): "Plea Bargaining and Social Welfare," *American Economic Review*, 73(4), 749–757.

Inderst, R. (2003): "Alternating-offer bargaining over menus under incomplete information," *Economic theory*, 22(2), 419–429.

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

Rabe, G., and D. Champion (2002): *Criminal Courts: Structure, Process, and Issues*. Prentice Hall.

Reinganum, J. (1988): "Plea Bargaining and Prosecutorial Discretion," *American Economic Review*, 78(4), 713–728.

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