Guilty Pleas by the Innocent*

Charles Z. Zheng†

July 2, 2025

Abstract

A criminal trial, modeled as an all-pay auction with strong correlation of types between the defendant and the prosecutor, fails to guarantee greater probabilities of conviction for the guilty than for the innocent, because the auction does not have any equilibrium where the defendant's distributional strategy is monotone. The criminal trial preceded by plea bargaining, by contrast, can guarantee that sometimes because under some parameter values the game has a separating equilibrium where only the guilty accepts the plea deal to avoid the trial. However, the existence of such separating equilibria is restricted by the incentive constraint that deters the prosecutor from making overly generous offers to pressure the innocent into guilty pleas. Such restriction is relaxed significantly if the boundary of admissible plea deals is set a priori by a neutral mediator, say the judge.

Keywords: All-pay auction, correlated types, plea bargain, criminal trial, pre-trial settlement, convicting the innocent

*The latest version is posted at this link. I thank Kathryn Spier for comments.

†Department of Economics, University of Western Ontario, London, ON, Canada, N6A 5C2, charles.zheng@uwo.ca. https://economics.uwo.ca/faculty/zheng/.

1 Introduction

This paper explores plea bargaining that precedes a criminal trial. The criminal trial follows an all-pay auction model where the two sides contest with their litigation efforts, each bearing the sunk cost of the effort, win or lose. Private information is stochastically correlated between the two sides. Plea bargaining takes the form of a take-it-or-leave offer made by the prosecutor for the defendant to respond. Without plea bargaining, the criminal trial fails to guarantee ex post larger conviction probabilities for the guilty than for the innocent. Plea bargaining improves such accuracy, but the improvement is restricted if the plea deal offer is unregulated. The restriction is relaxed if the offer is regulated by a cap and a floor that are derived from the primitives.

Most of the criminal cases in the United States are resolved by plea bargaining.[1] The sentencing reforms that mandate harsher or more lenient sentencing affect the discretion that a prosecutor can have during plea bargaining, which in turn impacts the outcome of most criminal cases. Debates on such reforms often center on questions how much and in which direction prosecutorial discretion in plea bargaining should be granted. There are claims in the literature suggesting that plea bargaining may hurt social welfare.[2] The theoretical question is whether plea bargaining improves social welfare at all and, if it does, in what format should plea bargaining take.

In the economic studies of plea bargaining, an earlier strand focuses on the cost-saving effect of plea bargaining (Adelstein [1] and Landes [21]). A later strand, initiated by Grossman and Katz [18], incorporates information asymmetry and focus on the screening value of plea bargaining. Most of this strand, Reinganum [25], Baker and Mezzetti [2] and Bjerk [7]), considers the strategic interaction between the defendant and the prosecutor during plea bargaining with the prosecutor assumed to have an objective approximate to social welfare. Recently, Siegel and Strulovici [30] characterize the optimal mechanism with the defendant being the only strategic player and explain plea bargaining as part of the optimal mechanism.

This paper stems from the second strand and departs from its predecessors by modeling

[1] The figure cited by Silveira [31] is over 90 percent, among federal cases and among state ones.

[2] Reinganum [25] presents a special case where the types of the defendant can be better separated if the prosecutor has no discretion at all during plea bargaining. Friedman and Wickelgren [16] suggest that plea bargaining hurts the accuracy of the legal system and generates a chilling effect on the legitimate activities that might be mistaken for harmful activities.

the criminal trial as a litigation game between the defendant and the prosecutor. In the predecessors, a criminal trial follows what Daughety and Reinganum [13] call the "perfect signal Bayesian court model": A signal is revealed to the court (or jury), which determines the outcome based on the signal and not on any action that the defendant or the prosecutor may take after the point where a trial is expected to occur;[3] and this signal is more likely to indicate guilt if the defendant is really guilty than if he is really innocent. Consequently, the trial by itself guarantees larger conviction probabilities for the guilty than for the innocent, and the screening value of plea bargaining relies on such accuracy condition of the trial.

By contrast, if a trial follows an all-pay auction as in this paper, the outcome is not determined by the signals privately known by the litigants, but rather by how the litigation efforts chosen by the two sides compete against each other in the court, and their efforts need not reflect the strength of their signals. Then the trial by itself fails to assure the accuracy condition. That is driven by an assumption that a prosecutor is more likely to get a strong signal for the guilt of the defendant when the defendant is really guilty than when he is really innocent (as in the predecessors' models). Consequently, in considering whether to match a high effort that may be exerted by the prosecutor, who is likely to have the strong signal if her prosecutorial effort reflects the strength of her signal, the guilty defendant believes it more likely than his innocent counterpart does that the high effort will actually be exerted by the prosecutor. Therefore, the guilty defendant gains more marginally than his innocent counterpart in matching such high efforts. But that means the guilty may exert more litigation effort than the innocent and hence, ex post, may get a larger acquittal probability than the innocent. This intuition is formalized as Proposition 1 in this paper that there exists no equilibrium of the auction game in which a guilty defendant almost surely exerts less litigation efforts than his innocent counterpart does.

Now that a criminal trial by itself fails to guarantee the accuracy of the legal system, plea bargaining could be understood as an indispensable instrument if, when it is added to precede the trial so that the trial can be avoided if a plea deal is reached, the game admits a (defendant-wise) *separating equilibrium* in which the guilty pleads guilty and the innocent goes to trial, for then the total probability of convicting the innocent, through pleas or trials, would never be larger than that of convicting the guilty. A somewhat unexpected lemma

[3]This signal may depend on the actions chosen by the players before the point that determines whether a trail will occur, as in Baker and Mezzetti [2] and Bjerk [7].

affords tractability to this game that would otherwise be complicated by the privately informed proposer (the prosecutor) and the endogenous interdependency between negotiation outcomes and continuation plays: In any separating equilibrium, the prosecutor offers the same plea deal regardless of her type (Lemma 2).[4] Then the incentive condition for the defendant in the equilibrium can be identified easily, which corresponds to an interval, uniquely determined by the parameters, containing all the payoffs offered to the defendant in any plea deals that the guilty is willing to accept and the innocent willing to reject. Consequently, the existence of separating equilibria, and hence the justification for plea bargaining, boil down to whether this exogenous interval is nonempty and whether the prosecutor can be kept from making offers outside the interval.

The finding via this approach is that the extent to which plea bargaining can improve the accuracy of the legal system is limited if it is unregulated and the limitation is much removed if plea bargaining is regulated. The limitation of unregulated plea bargaining is driven by the bidding equilibrium in the criminal trial that ensues after the prosecutor makes an offer and the innocent defendant rejects it according to the supposedly separating equilibrium. In this bidding equilibrium, either the innocent defendant or the weak prosecutor gets zero surplus from the trial. In the former case, the innocent defendant would rather plead guilty than go through the costly trial, namely, the aforementioned interval is empty. In the latter case, the weak prosecutor is tempted to avoid the costly trial by deviating outside the interval with an offer so generous that the innocent defendant does not reject no matter what continuation equilibrium he may have in mind. Thus, separating equilibria do not exist when the prosecutor is sufficiently resourceful or when her weak type is sufficiently informative about the innocence of the defendant (Proposition 2). Meanwhile, separating equilibria do exist under the parameter values such that the defendant is significantly more resourceful than the prosecutor and when the prosecutor's weak type does not imply too large a probability of the defendant's innocence (Proposition 3). Thus, plea bargain does improve the accuracy of the legal system sometimes, though the extent of the improvement is quite restricted.

[4]Although Myerson's [24] inscrutability principle implies, in the mediated communication game equivalence of the plea bargain game, that it is without loss to assume that the prosecutor discloses no information at her initial stage of communication, the principle does not imply that the outcome of the communication game in the event where plea deals are accepted is unconditional on the prosecutor's type. Lemma 2, by contrast, implies that the said outcome is unconditional on the prosecutor's type.

However, the analysis naturally leads to a way to weaken this restriction. If the parameter values are such that the aforementioned exogenous interval is empty, then there is no hope to have separating equilibria anyway. Else, the said interval is nonempty, then the difficulty to have separating equilibria comes purely from the incentive constraint that keeps the prosecutor from all sorts of deviations, each of which needs to be deterred by some continuation equilibrium that both parties need to be incentivized to play. But this difficulty can be easily done away with by having a neutral mediator, say the judge, set up a priori the ceiling and floor for the payoffs that the prosecutor is allowed to offer the defendant in any plea deal. Then the deviations of offering plea deals that go beyond the ceiling and floor become unavailable to the prosecutor. When the ceiling and floor are defined to coincide with the aforementioned interval, the only deviation that requires deterrence is the prosecutor's offering no plea deal at all. That weakens the incentive constraint for the prosecutor significantly. Consequently, separating equilibria obtain, and hence guarantees the accuracy of the legal system, under a much weaker condition when the prosecutor's set of admissible offers is regulated by an interval determined by the primitives (Proposition 4).

In light of Proposition 4, the regulation in practice that requires plea deals be approved by the judge[5] can be understood as an institutional attempt to keep the prosecutor from offering overly generous or overly draconian deals that pressure the innocent into guilty pleas. The only further condition that Proposition 4 requires is that the boundary of admissible plea deals be either established, or commonly expected, before the plea bargaining stage.

Proposition 4 sheds some light on the effect of the policies that mandate harsher sentencing (cf. Silveira [31]). Mandating harsher sentencing amounts to lowering the ceiling of the payoffs that the prosecutor is allowed to offer in plea deals. If the original ceiling before such a policy was nonexistent or above the exogenous ceiling in the aforementioned interval, the policy has the positive effect of making some upward deviations unavailable to the prosecutor thereby improving the prospect of separating equilibria. If the original ceiling is already within the said interval, by contrast, the policy has no positive effect. Worse yet, if the mandated sentencing is so harsh that it pushes the ceiling below the floor of the exogenous interval, then the policy annihilates any separating equilibrium, as no plea deal given the overly harsh sentencing mandate is acceptable to even the guilty defendant.

[5]A widely reported recent case is that a judge's disapproved the plea deal between Hunter Biden and the United States Department of Justice prosecutor in 2023.

Reinganum [25] has a special case where a separating equilibrium exists if the prosecutor is restricted, by an exogenous policy, to pooling strategies that make the same offer regardless of her type. My paper implies that there is no need to require such a restriction by a policy, because in any separating equilibrium the prosecutor necessarily chooses a pooling strategy. This difference leads to opposite policy implications. Reinganum's pooling-strategy policy cannot be implemented unless it becomes a mandate of a specific plea deal that the prosecutor is required to offer. Otherwise, on the equilibrium path only one offer is observed, and so it is impossible to verify whether the prosecutor has played a pooling strategy. Thus, unless the policy maker knows exactly what the equilibrium plea deal is, the mandate would ruin the separating equilibrium. My finding, by contrast, is that the policy maker needs only to define the boundary within which the prosecutor enjoys full discretion, and that the boundary can be calculated from the primitives.

All-pay auction has been used to model conflict in a recent literature of conflict preemption such as Balzer and Schneider [3], Kamranzadeh and Zheng [20], Lu, Lu and Riis [22], Schouten [28], and Zheng [33, 34].[6] The main focus of that literature is the possibility for all types to settle thereby avoiding the conflict.[7] This paper, by contrast, focuses on the possibility to separate the two types of the player under consideration so that only the weaker type settles. Other than the all-pay auction model, Tullock contests have been used to model civil trials by Rosnberg and Spier [27] on the incentive for litigation investment in the context of class actions, and Chen and Wang [10] on fee shifting rules in litigation.[8]

The paper also adds to the all-pay auction literature with the observation that given strong correlation between the bidders the game admits no equilibrium, symmetric or asymmetric, in which some bidder's strategy is monotone. The strong correlation assumption is opposite to Siegel's [29] condition that guarantees both bidders' strategies to be monotone in

[6]The earlier works in that field include Bester and Wärneryd [6], Compte and Jehiel [11], Fey and Ramsay [15], Hörner, Morelli and Squintani [19], and Spier [32].

[7]Except Kamranzadeh and Zheng [20], who focus on the total ex ante expected payoff of the contestants.

[8]Considering a settlement game before a civil trial, Chen and Wang [10] model the trial as a special case of the perfect signal Bayesian court in which the court will know perfectly whether the defendant is liable or not. Thus the trial guarantees accuracy by assumption, and hence the settlement negotiation does not add any more in that regard, though it works as a screening device because the authors also assume that a liable defendant pays the damage if and only if he loses in the trial and that the probability of losing the trial is a function of the litigation efforts exerted by the two sides. Expecting to incur a large effort cost that is needed to prevail in the trial, the liable defendant is more willing to avoid it through settling.

equilibrium. Without Siegel's condition, Rentschler and Turocy [26] provide various examples in which non-monotone equilibria exist. Schouten [28] shows nonexistence of symmetric monotone equilibria in a symmetric model based on strong correlation. Bedard and Zheng [5] observe nonexistence of monotone equilibrium in a special case of the all-pay auction in this paper where the defendant's marginal cost of bids is a commonly known constant.

2 Criminal Trials as All-Pay Auctions

A criminal trial is all-pay in the sense that each side, win or lose, has to bear the sunk cost for the resources that it devotes to the trial. The information asymmetry feature in auctions is also salient in criminal trials: The defendant knows privately whether he is guilty or innocent, and the prosecutor has better knowledge about the evidence against the defendant than the latter does. This information asymmetry, different than that in most textbook examples in auction theory, is complicated by the stochastic correlation between the two sides' private information: The guilty defendant with more likelihood believes that the prosecutor's case against him is strong than the innocent defendant does, other things equal. And likewise the prosecutor with a strong case against the defendant has more confidence that the defendant is guilty than the prosecutor with a weak case does. Let us therefore model a criminal trial as the following all-pay auction game between the prosecutor and the defendant whose types (private information) are correlated to each other.

The game has two players, d (defendant) and p (prosecutor). Player d's type t_d is either g (guilty) or i (innocent), and p's type t_p is either s (strong evidence against d) or w (weak evidence against d), each privately known to the player. Assume $g, i, s, w \in \mathbb{R}$ and

$$s > w > 0 \quad \text{and} \quad i > g > 0. \tag{1}$$

For each player $j \in \{d, p\}$, let $f_j(t_j \mid t_{-j})$ denote the probability that j's type is t_j conditional on the rival $-j$'s type being t_{-j}. Types are correlated between the players in the sense that

$$f(g|s) > f(g|w) \quad \text{and} \quad f(w|i) > f(w|g), \tag{2}$$

where the subscripts d and p in $f_d(g|s)$, $f_d(g|w)$, $f_p(w|i)$ and $f_p(w|g)$ are suppressed because types are represented by distinct symbols across players. Nonetheless, this model allows for asymmetry between the two players, as the belief $f_d := (f(\cdot|s), f(\cdot|w))$ about the defendant d

can be different from the belief $f_p := (f(\cdot|i), f(\cdot|g))$ about the prosecutor p, and so are their type spaces $T_d := \{i, g\}$ versus $T_p := \{s, w\}$.

Each player j privately knows his or her own type t_j, then obtains the belief $f(\cdot|t_j)$ about the opponent $-j$'s type, and then submits a sealed bid $b_j \in \mathbb{R}_+$ (total amount of litigation efforts and resources committed at the outset) and bears its sunk cost, assumed equal to b_j/t_j. The higher bidder is the winner, with ties (equal bids) broken randomly with equal probabilities. If d is the winner then the defendant is acquitted, else p is the winner and the defendant is convicted. The payoff for each player $j \in \{d, p\}$ of type t_j is equal to $\overline{v}_j - b_j/t_j$ if j is the winner, and otherwise equal to $\underline{v}_j - b_j/t_j$, where $\underline{v}_j < 0 < \overline{v}_j$ $(\forall j \in \{d, p\})$ are commonly known parameter.[9] For instance, if $j = d$ (defendant) and d does not win, then the defendant is convicted and his payoff (or rather penalty) equals $\underline{v}_d < 0$.

Let $\mathbf{1}_j$ denote the indicator function for the event that player j wins $(\forall j \in \{d, p\})$. Then the payoff for player j $(\forall j \in \{d, p\})$ of type t_j after bidding b_j is equal to

$$\overline{v}_j \mathbf{1}_j + \underline{v}_j (1 - \mathbf{1}_j) - b_j/t_j = \underline{v}_j + \left(\overline{v}_j - \underline{v}_j\right) \mathbf{1}_j - b_j/t_j = \underline{v}_j + v_j \mathbf{1}_j - b_j/t_j,$$

with $v_j := \overline{v}_j - \underline{v}_j$. Thus, without loss, we can reduce the parameters $\overline{v}_j$ and $\underline{v}_j$ to v_j. Since $v_j > 0$, the above-displayed payoff for player-type (j, t_j) is equivalent to $\mathbf{1}_j - b_j/(v_j t_j)$. Scale the original values of t_j by v_j and denote the new values thereby obtained by the same notation t_j. Obviously (1) is preserved. For the rest of the paper, therefore, let us assume without loss that the payoff for player-type (j, t_j) that has bid b_j is equal to

$$\mathbf{1}_j - \frac{b_j}{t_j},$$

where $t_j \in T_j$ ($T_d = \{i, g\}$ and $T_p = \{s, w\}$).

Throughout the paper, the prior beliefs are assumed to be *nondegenerate*. That is, $f(t_j|t_{-j}) > 0$ for each pair (t_j, t_{-j}) of types across the two players.

[9] The selfishness assumption about the prosecutor, consistent with the constitutional framers' basic notion about the corruptibility of individuals in power, departs from much of the theoretical literature on plea bargaining that assumes the prosecutor's objective to approximate social welfare (Grossman and Katz [18], Reinganum [25], Baker and Mezzetti [2] and Bjerk [7]). Some other works in that field, Landes [21] and Daughety and Reinganum [12, 14], do assume or include selfish prosecutors. Empirical evidence for prosecutors being selfish is provided by Boylan [8], Boylan and Long [9], McCannon [23] and Bandyopadhyay and McCannon [4].

3 The Drive for Plea Bargaining

A criminal justice system can be viewed as a communication game, possibly multistage, that results in two outcomes regarding the defendant, either his being convicted by trial or by a guilty plea, or his being acquitted by trial.[10] As a communication game, the system can be evaluated according to an associated equilibrium, with the equilibrium concept being perfect Bayesian equilibrium (PBE) if the game is multistage, or Bayesian Nash equilibrium (BNE) if it is single stage such as the all-pay auction modeled before.

Criminal justice systems in the real world are known to be susceptible to errors, acquitting the guilty or convicting the innocent. It is natural to require that the guilty should be weakly less likely to be acquitted than the innocent, or equivalently, the innocent be weakly less likely to be convicted than the guilty.[11] This is formulated as the following normative condition regarding an equilibrium $\mathscr{E}$ of the criminal justice system under consideration:

> Conditional on any set of action profiles (across players) that occurs with a positive probability on the path of $\mathscr{E}$, the guilty defendant's probability of being acquitted is less than or equal to the innocent defendant's probability of being acquitted. (A)

For example, in Grossman and Katz's [18] model, a criminal trial is assumed to be a lottery that results in conviction with an exogenous probability that is larger for the guilty than for the innocent. The prosecutor, assumed to have no private information, proposes a plea deal as a take-it-or-leave offer. In equilibrium, therefore, the defendant self-selects in response to offer, accepting it if he is guilty and rejecting it to let the chips fall in the

[10] Gross et al. [17] have an estimate of false convictions among death sentences in the United States.

[11] The accuracy of the criminal justice system has been suggested as a normative condition by Grossman and Katz [18]. Friedman and Wickelgren [16] underscores the importance of the accuracy condition by suggesting that the inaccuracy of the criminal justice system generates a chilling effect on the legitimate activities that might be mistaken for harmful activities.

When a criminal trial is based on the perfect signal Bayesian court model (Grossman and Katz [18], Reinganum [25], etc.), it makes no difference whether the condition is required ex ante or ex post (after the litigation efforts of the opponent have been chosen). The accuracy condition formulated here takes the ex post standpoint, which is in line with the ex post implementation perspective. For instance, in the situations suggested by Friedman and Wickelgren [16], an individual deciding whether to undertake a legitimate activity that might be mistaken for harmful ones (e.g., an innovative surgery) may feel ambiguous about the belief that a potential prosecutor may have about his liability should he undertake the activity and get indicted. Being belief-free, the ex post accuracy condition would be appropriate to such situations.

trial if he is innocent. Thus their model satisfies the normative condition (A), "A" for the "accuracy of the legal system" suggested by Grossman and Katz. Reinganum [25] makes essentially the same assumption about the conviction probability in the lottery-like trial,[12] with the probability of conviction assumed privately known to the prosecutor. At equilibrium, the prosecutor's offer reflects her type partially, the larger the conviction probability she knows, the harsher the plea deal she offers. The defendant updates about his probability of conviction and then responds to the offer. Thus, the guilty is more likely to accept a plea deal than the innocent is, and conditional on going to trial, the guilty is more likely to be convicted than the innocent. Hence again Condition (A) is satisfied.

In my model of criminal trials, by contrast, the probability of conviction is not exogenous but rather depends on the actions that the two opponents take during the trial. Consequently, the assumption about the conviction probability in the previous models becomes an assertion about equilibria. The rest of this section shows that this assertion is false and hence a criminal trial by itself fails the normative condition (A).

A criminal trial that ensues without a plea bargaining stage corresponds to a single-stage communication game, the all-pay auction defined before with the belief system (f_d, f_p) held at the start of the auction, and so BNE is the equilibrium concept. For an BNE of the criminal trial to satisfy Condition (A), the next lemma notes that the defendant's equilibrium strategy needs to be monotone in the sense that the bids submitted by the innocent defendant are almost surely higher than the bids submitted by the guilty defendant.

Formally, let $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ denote any BNE of the auction so that $\phi_j^{t_j}$ denotes the cumulative distribution function (cdf) according to which the player-type (j, t_j) (player j of type t_j) chooses a bid to submit. For each player-type (j, t_j) denote the support of $\phi_j^{t_j}$ by $B_j^{t_j}$. A player j's strategy $(\phi_j^{t_j})_{t_j \in T_j}$ is said to be *monotone* iff $t_j > t_j' \Rightarrow \inf B_j^{t_j} \geq \sup B_j^{t_j'}$.

Lemma 1 *For any BNE of the all-pay auction, if it satisfies Condition (A), then the defendant's strategy in this BNE is monotone.*

Lemma 1 is due to an observation (Appendix A.1) that in any BNE of the auction

[12] In Reinganum's [25] model, the probability of conviction is randomly drawn according to an exogenous distribution conditional on the defendant's type, and Grossman and Katz's assumption that the lottery results in a larger conviction probability for the guilty than for the innocent is expressed in a conditional expectation form (conditional on any possible set of values that the probability may take).

game, each player's equilibrium bid distribution has no gap, namely,

$$B_d^i \cup B_d^g = B_p^s \cup B_p^w = \left[0, \overline{b}\right] \tag{3}$$

for some $\overline{b} > 0$ associated with the BNE.[13] Consequently, if the defendant picks a low bid b when he is innocent and a high bid b' when he is guilty, which occurs with positive probability unless his strategy is monotone, then the probability with which his opponent p bids between b and b' is positive due to the no-gap observation. In this event, the innocent loses while the guilty wins, and thus Condition (A) is not met.[14] □

A downside of criminal trials without plea bargaining is signified by the next proposition. It says that the all-pay auction admits no equilibrium in which the defendant's strategy is monotone and hence, by Lemma 1, no equilibrium to satisfy Condition (A), when the belief system is strongly correlated. A belief system (f_d, f_p) is said to be *strongly correlated* iff

$$sf(i|s) < wf(i|w) \quad \text{and} \quad if(s|i) < gf(s|g). \tag{4}$$

Condition (4) captures strong correlation between the two players' types because it implies the weaker correlation condition (2), as $s > w > 0$ and $i > g > 0$. Although it is known in the literature that (4) implies existence of BNEs in some cases where neither bidders' equilibrium strategies are monotone;[15] the following proposition is new, observing that (4) precludes existence of any BNE, symmetric or asymmetric, in which at least one bidder's strategy is monotone.[16]

Proposition 1 *If the belief system (f_d, f_p) in the all-pay auction game is strongly correlated, then there exists no BNE in which the defendant's strategy is monotone and hence there exists no BNE that satisfies Condition (A).*

To sketch the reasoning for Proposition 1 (proved in Appendix A.2), consider any BNE $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ of the auction so that $\phi_j^{t_j}$ denotes the cdf according to which player-type (j, t_j)

[13] However, the bidding distribution $\phi_j^{t_j}$ for a player-type (j, t_j) could have gaps, namely, $B_j^{t_j}$ could be disconnected. A gap of B_p^s, say, can be filled up by B_p^w because monotonicity of (ϕ_p^s, ϕ_p^w) is not assumed.

[14] Conversely, if a BNE of the all-pay auction is monotone and symmetric, with symmetry implying zero probability of ties in the equilibrium, one readily sees that the BNE satisfies Condition (A).

[15] The opposite of (4) is proposed by Siegel [29] to guarantee monotonicity of BNEs.

[16] While the proposition is a statement about the defendant, the statement that obtains through switching the roles between the players can use the same proof that switches the roles accordingly.

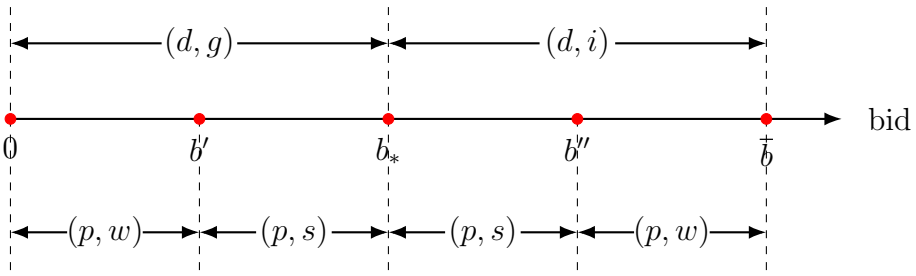

Figure 1: Monotonicity of ϕ_d implies ϕ_p is non-monotone above b_* and monotone below b_*

chooses a bid to submit, with the support of $\phi_j^{t_j}$ denoted by $B_j^{t_j}$. As mentioned before, Eq. (3) holds. That is, $B_d^i \cup B_d^g$ and $B_p^s \cup B_p^w$ are equal to a nondegenerate interval $[0, \bar{b}]$.

Suppose, to the contrary of Proposition 1, that the defendant's equilibrium strategy (ϕ_d^i, ϕ_d^g) is monotone. Then B_d^i is an upper subinterval $[b_*, \bar{b}]$ of the aforementioned $[0, \bar{b}]$, and B_d^g the lower subinterval $[0, b_*]$ of $[0, \bar{b}]$. Figure 1 shows the implication: If both types of the prosecutor p participate in bidding above b_* (against the innocent defendant (d, i)), the weak type (p, w) would bid *higher* than the strong type (p, s). If both types of p participate in bidding below b_* (against (d, g)), by contrast, the weak type would bid lower than the strong type.

The non-monotonicity of the prosecutor's strategy (ϕ_p^s, ϕ_p^w) on $(b_*, \bar{b}]$ is driven by the strong correlation condition, Ineq. (4): Since the weak prosecutor assigns a much larger probability than the strong prosecutor does to the event that the defendant is innocent, the weak prosecutor's "marginal revenue"—the derivative $f(i|w)\dot{\phi}_d^i$ of the opponent d's bid distribution conditional on her own weak type—exceeds her marginal cost $1/w$ by a larger quantity than the strong prosecutor's marginal revenue $f(i|s)\dot{\phi}_d^i$ does her marginal cost $1/s$.[17] Thus the weak prosecutor bids more aggressively than the strong does if both bid above b_*.

Analogously, in bidding below b_* the prosecutor is competing against the guilty type g. Then, since $f(g|s) > f(g|w)$ by the correlation condition (2), it is the strong type s of the prosecutor that assigns a larger probability to the event of running into a competing bid below b_*, and hence has a larger marginal revenue from such bids. This, coupled with the

[17] Let $\dot{\varphi}$ denote the derivative $\frac{\partial}{\partial b}\varphi$ if φ is a.e. differentiable. Although the usual no-atom and no-gap arguments are inapplicable here due to type correlation between the two players and possible asymmetry of their strategies, the monotonicity supposition $\sup B_d^g \leq \inf B_d^i$, coupled with the no-gap observation Eq. (3), allows one to prove that none of ϕ_d^i, ϕ_p^s and ϕ_p^w has any atom (mass point) in $(b_*, \bar{b}]$.

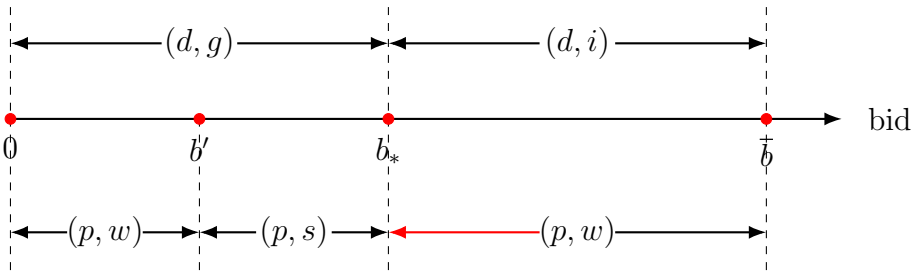

Figure 2: To keep (d, g) from deviating above b_*, (p, s) does not bid above b_*

fact that the strong prosecutor incurs less marginal cost than the weak one, implies that the strong type bids more aggressively than the weak type regarding bids below b_*. That is why (ϕ_p^s, ϕ_p^w) is monotone on $[0, b_*)$.

Figure 2 shows the consequence of Figure 1: The guilty defendant would deviate from submitting only the bids below b_* to submitting the bids in (b_*, b'') against the strong prosecutor as long as the latter bids there at all. In making such a deviating increase of bids, the guilty defendant's marginal revenue is the derivative $f(s|g)\dot{\phi}_p^s$ of the opponent's bid distribution on (b_*, b'') conditional on his type being guilty. To rationalize any bid in (b_*, b'') for the innocent defendant (d, i), we need $\dot{\phi}_p^s = 1/(if(s|i))$ on (b_*, b''). Thus, the marginal revenue of the guilty defendant's deviating bid increase is equal to $f(s|g)/(if(s|i))$, which by the strong correlation condition (4) is greater than his marginal cost $1/g$. Hence he would deviate unless, as in Figure 2, (p, s) does not bid above b_* at all.

Furthermore, the strong prosecutor does not bid in $(0, b_*)$ either, as depicted in Figure 3. The reasoning is similar to that in the previous paragraph.[18] The weak prosecutor (p, w), who is supposed to bid above b_* and not bid within (b', b_*) according to Figures 1 and 2, would deviate to the bids in (b', b_*) unless $(b', b_*) = \varnothing$.

In sum, the strong prosecutor is not supposed to submit any bid other than zero and hence she gets only zero surplus in the equilibrium. Then she would deviate to some positive bids thereby getting a positive surplus. This contradiction establishes the impossibility for the defendant to have a monotone equilibrium strategy. □

In light of Proposition 1, plea bargains, which are often observed in practice, can be understood as attempts to improve the prospect for a criminal justice system to satisfy

[18] From the fact that the support of ϕ_p^s is restricted to $[0, b_*]$ according to Figure 2, one can show that $(0, b_*]$ is nondegenerate and contains no atom for ϕ_d^i, ϕ_p^s, or ϕ_p^w.

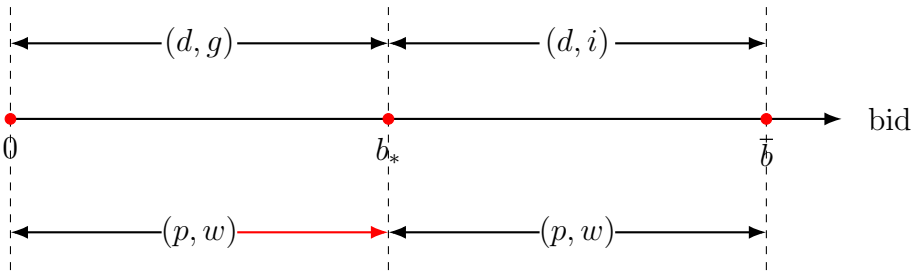

Figure 3: (p, w) would deviate unless (p, s) does not bid in $(0, b_*]$

Condition (A). If somehow the prosecutor is willing to offer a plea deal that is accepted only by the guilty, so that only the innocent goes to the trial and stands a chance to prevail, then the guilty is never "found" innocent, and the innocent may be acquitted. The question is: Are such improvements warranted at equilibrium?

4 The Plea Bargain Game

Let us consider a multistage game where plea bargaining may preempt the criminal trial. First, each player $j \in \{d, p\}$ is privately informed of his or her type t_j. Second, the prosecutor p chooses an $x \in [0, 1]$. If $x = 0$ then no plea deal is offered and the criminal trial (the all-pay auction in Section 2) proceeds. Else $(x > 0)$ then the prosecutor offers to the defendant d a plea deal of paying d an amount equal to x if d pleads guilty without a trial. Then the defendant chooses whether to accept the offer. If he accepts it then the game ends with the defendant having a payoff equal to x and the prosecutor having a payoff equal to $1 - x$ (1 being the utility of winning without cost, given the payoff normalization in Section 2).[19] If the defendant rejects the offer then the criminal trial ensues.[20]

The equilibrium concept natural to this multistage game is perfect Bayesian equilibrium

[19] Without the payoff normalization, a plea deal offer x means that if the defendant pleads guilty according to the deal then his payoff increases from the (negative) payoff $\underline{v}_d$ of being convicted to $\underline{v}_d + (\overline{v}_d - \underline{v}_d)\, x$, so x captures the relative reduction of sentencing.

[20] As in the previous plea bargaining literature, I make the commitment assumption that the trial ensues once the plea deal is rejected even if only the innocent rejects the deal. The assumption is regarded to be realistic in the recent empirical work by Silveira [31]. Siegel and Strulovici [30] point to various actual practices in the legal system as justification for the assumption.

(PBE). Denote any PBE of the game by

$$\left((\sigma_p^s, \sigma_p^w), (\sigma_d^{i,x}, \sigma_d^{g,x}, \hat{f}_p^x, \hat{f}_d^x, \phi_p^{s,x}, \phi_p^{w,x}, \phi_d^{i,x}, \phi_d^{g,x})_{x\in[0,1]}\right)$$

such that $\sigma_p^{t_p}$ represents the cdf according to which the prosecutor p of type t_p chooses an $x \in [0,1]$, $\sigma_d^{t_d,x}$ represents the probability with which the defendant d of type-t_d accepts x ($\sigma_d^{t_d,0} = 0$, namely, the defendant "chooses" to reject the null offer $x = 0$ for sure, though the null offer launches the criminal trial ensues without his response), $\hat{f}_p^x := (\hat{f}_p^x(\cdot|i), \hat{f}_p^x(\cdot|g))$ denotes the posterior belief about the prosecutor's type conditional on her having chosen x, $\hat{f}_d^x := (\hat{f}_d^x(\cdot|s), \hat{f}_d^x(\cdot|w))$ the posterior belief about the defendant's type conditional on his having rejected the offer x, and $(\phi_p^{s,x}, \phi_p^{w,x}, \phi_d^{i,x}, \phi_d^{g,x})$ the BNE, associated with this PBE, of the all-pay auction given posterior belief system $(\hat{f}_p^x, \hat{f}_d^x)$ conditional on x having been offered by p and rejected by d.

A PBE of the plea bargaining game is said to be *separating* iff on the equilibrium path each type of the prosecutor offers some non-null plea deal, the guilty defendant accepts the offer for sure, and the innocent defendant rejects it for sure. That is, $\sigma_p^{t_p}(0) = 0$ for each $t_p \in \{s, w\}$, and $\sigma_d^{g,x} = 1$ and $\sigma_d^{i,x} = 0$ for any nonzero x in the support of σ_p^s or that of σ_p^w.

Obviously, any separating PBE of the plea bargaining game satisfies Condition (A), as the guilty defendant, always accepting the on-path plea deal to plead guilty, is never acquitted. Let us restrict attention to separating PBEs for the normative appeal, as well as the tractability. The question is: To what extent do separating equilibria exist?

5 The Condition for Separating Plea Bargains

To examine the possibility of separating PBEs in the plea bargain game, let us start with the continuation game on the equilibrium path at the event where a plea deal has been offered and rejected, and hence the criminal trial unfolds as the all-pay auction modeled before. Since the equilibrium is separating, the defendant's having rejected the offer reveals that his type is innocent. Denote this degenerate belief by $\hat{f}_{d=i}$, namely, $\hat{f}_{d=i}(i|t_p) = 1$ conditional on each type $t_p \in \{s, w\}$ of the prosecutor. Denote the posterior belief about the prosecutor's type, conditional on her having made that offer on path, by $\hat{f}_p$ $(= (\hat{f}(\cdot|i), \hat{f}(\cdot|g)))$, with the superscript that indicates the offer suppressed. Thus, the continuation game being considered is the all-pay auction with belief system $(\hat{f}_{d=i}, \hat{f}_p)$. For each player-type (j, t_j),

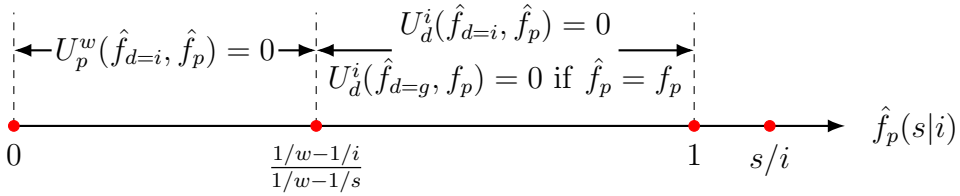

Figure 4: The case of $s/i \geq 1$

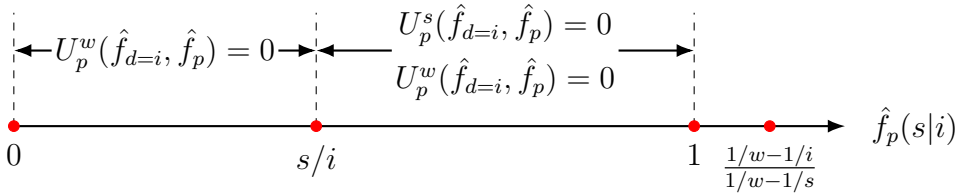

Figure 5: The case of $s/i \leq 1$

let $U_j^{t_j}(\hat{f}_{d=i}, \hat{f}_p)$ denote the expected payoff that (j, t_j) gets from best-responding to the BNE of the all-pay auction that is defined by the belief system $(\hat{f}_{d=i}, \hat{f}_p)$, or loosely called the "surplus" from the said BNE.

It turns out that the BNE of this continuation game, and each player-type's surplus therein, are determined by the value of the posterior probability $\hat{f}_p(s|i)$ relative to the parameters s, w and i. This is shown in Figure 4 for the case $s/i \geq 1$ and Figure 5 for the case $s/i \leq 1$ (proved in Appendix A.3). The parametric cutoffs s/i and $(1/w - 1/i)/(1/w - 1/s)$ are positioned differently in the two figures because

$$s \geq (>)\ i \iff \frac{s}{i} \geq (>)\ 1 \geq (>)\ \frac{1/w - 1/i}{1/w - 1/s}. \tag{5}$$

Note from the figures that either the innocent defendant gets zero surplus (when $(1/w - 1/i)/(1/w - 1/s) \leq \hat{f}_p(s|i) \leq 1$) or the weak prosecutor gets zero surplus from the BNE. This fact leads to the next lemma, which is somewhat unexpected, observing that the prosecutor necessarily pools her types in any PBE that separates the guilty from the innocent.

Lemma 2 *In any separating PBE of the plea bargain game there is a unique $x \in (0, 1]$ such that both types of the prosecutor offer x for sure as the plea deal.*

To prove the lemma, consider any separating PBE. On its equilibrium path, by the definition of separating PBE, each type of the prosecutor always makes some positive offer,

which the guilty defendant accepts for sure and the innocent defendant rejects for sure. Thus, in the event where any such on-path offer x (> 0) is proposed and rejected, the all-pay auction ensues with belief system $(\hat{f}_{d=i}, \hat{f}_p^x)$, where $\hat{f}_p^x$ denotes the posterior belief about the prosecutor's type conditional on her having offered x. By the aforementioned fact shown in Figures 4 and 5, either (i) $U_d^i(\hat{f}_{d=i}, \hat{f}_p^x) = 0$ or (ii) $U_p^w(\hat{f}_{d=i}, \hat{f}_p^x) = 0$. Case (i) is impossible in the separating PBE, because the innocent defendant would accept any positive offer to avoid the criminal trial that gives him zero surplus. Thus, the only possibility is Case (ii), where the weak prosecutor expects zero surplus from the criminal trial that may occur on path given any offer x that she may propose according to the equilibrium. Consequently, in making any such on-path offer x, her expected payoff is equal to $(1-x)f(g|w)$ because in the separating PBE the defendant accepts x if and only if he is guilty, which occurs with probability $f(g|w)$ according to the prosecutor's prior belief conditional on her weak type. Since $f(g|w) > 0$ (the nondegenerate prior belief assumption), it follows that in the best response for (p, w), only the minimum among all these x is offered. That is, there exists a unique $x_* > 0$ that (p, w) offers for sure on the equilibrium path.

It follows that on the equilibrium path the strong prosecutor also offers this x_* for sure. Suppose, to the contrary, that she may offer some $x \neq x_*$ on path. Then her type is revealed and the posterior about her type is $\hat{f}_p^x(s|i) = 1$. By the fact shown in Figures 4 and 5, and the fact that we are currently in the case where $U_d^i(\hat{f}_{d=i}, \hat{f}_p^x) \neq 0$, one can show that her surplus $U_p^s(\hat{f}_{d=i}, \hat{f}_p^x)$ given this posterior ($\hat{f}_p^x(s|i) = 1$) is equal to zero.[21] Then the expected payoff for (p, s), in offering such x, is equal to $(1-x)f(g|s)$, or equivalently $1-x$ modulo the positive parametric scalar $f(g|s)$. Thus, just as in the case for (p, w), there exists a unique positive $x^* \neq x_*$ that (p, s) offers on the equilibrium path other than x_*. If $x_* \neq x^*$ then either type of the prosecutor would deviate to $\min\{x_*, x^*\}$. Thus, $x_* = x^*$, which proves Lemma 2. □

Thanks to Lemma 2, in any separating PBE there exists an $x_* \in (0, 1]$ that each type of the prosecutor offers for sure as the plea deal. If x_* is offered, then there is no updating about the prosecutor, and the defendant chooses his response by comparing the offer x_* to his surplus from the criminal trial that ensues if he rejects the offer. In the event of the

[21] From Figures 4 and 5 we see that $U_d^i(\hat{f}_{d=i}, \hat{f}_p^x) \neq 0$ implies that it is impossible to have $(1/w - 1/i)/(1/w - 1/s) \leq \hat{f}_p^x(s|i) = 1$. Then (5) implies $s < i$ and hence $s/i < 1$ as in Figure 5. This coupled with $\hat{f}_p^x(s|i) = 1$ implies that the right interval in Figure 5 is the only possibility.

criminal trial, since the offer x_* is on-path and the PBE is separating, the posterior system is $(\hat{f}_{d=i}, f_p)$, f_p being the prior about t_p. Thus, the incentive condition for the defendant is

$$U_d^i(\hat{f}_{d=i}, f_p) \geq x_* \geq U_d^g(\hat{f}_{d=i}, f_p), \tag{6}$$

where the first inequality says rejecting x_* is a best response for the innocent defendant, and the second inequality says accepting x_* is a best response for the guilty defendant.

The incentive condition for the prosecutor is more complicated. For any $x \neq x_*$, let $\left(\sigma_d^{i,x}, \sigma_d^{g,x}, \hat{f}_p^x, \hat{f}_d^x, \phi_p^{s,x}, \phi_p^{w,x}, \phi_d^{i,x}, \phi_d^{g,x}\right)$ denote a PBE of the continuation game starting from the event that x has just been offered. Here $(\sigma_d^{i,x}, \sigma_d^{g,x})$ represents the defendant's (type-dependent) response to the offer x, $\hat{f}_p^x$ represents the off-path posterior about the prosecutor's type, $\hat{f}_d^x$ the posterior belief about the defendant's type, which is off-path if rejecting x is not played with a positive probability according to $(\phi_d^{i,x}, \phi_d^{g,x})$. And $(\phi_p^{s,x}, \phi_p^{w,x}, \phi_d^{i,x}, \phi_d^{g,x})$ denotes a BNE of the criminal trial with posterior system $(\hat{f}_p^x, \hat{f}_d^x)$. Denote the type-$t_p$ prosecutor's expected payoff from this PBE of the continuation game by $\tilde{U}_p^{t_p}(x)$. Then the incentive condition for the prosecutor is

$$\forall x \in [0,1] \setminus \{x_*\}\ \forall t_p \in \{s,w\} : f(g|t_p)(1-x_*) + f(i|t_p)U_p^{t_p}(\hat{f}_{d=i}, f_p) \geq \tilde{U}_p^{t_p}(x). \tag{7}$$

In sum, a separating PBE of the plea bargain game exists if and only if there exists an $x_* \in (0,1]$ that satisfies both (6) and (7).

6 Guilty Pleas by the Innocent

From the necessary and sufficient condition for separating PBEs, which consists of (6) and (7) in the previous section, we can see that the prospect for the plea bargain game to have separating PBEs is somewhat meager. Once the plea bargain stage is added to a criminal trial, it is hard to keep the innocent defendant from pleading guilty to avoid the costly trial.

This problem is due to the performance of the criminal trial that will occur if the prosecutor offers plea deals according to the supposedly separating PBE and is rejected when the defendant happens to be innocent. In that event, as shown in Figures 4 and 5, either (i) $U_d^i(\hat{f}_{d=i}, f_p) = 0$ or (ii) $U_p^w(\hat{f}_{d=i}, f_p) = 0$. We have already seen, in the reasoning for Lemma 2, that separating PBEs do not exist in Case (i), where the innocent defendant would take the on-path plea deal to avoid the trial, as the trial can only give him zero surplus. That is, (6) cannot hold because the upper bound there equals zero while $x_* > 0$.

In Case (ii), (7) may be violated because the weak prosecutor (p, w) may deviate to making an offer so large that even the innocent defendant would accept and plead guilty: The weak prosecutor in Case (ii) expects only zero surplus if she makes the on-path offer x_* and the defendant happens to be innocent. By contrast, if she makes an offer $x > x_*$ that is acceptable to the innocent defendant, the weak prosecutor would get a positive surplus $1 - x$ if the defendant is innocent. This deviation would reduce her payoff (from $1 - x_*$ to $1 - x$) if the defendant turns out to be guilty. However, the probability $f(g|w)$ of the latter event is relatively small conditional on her type being weak, as the weak evidence against the defendant makes her unsure of the defendant's guilt. The more informative is the prosecutor's evidence, the smaller is $f(g|w)$ compared to the probability $f(i|w)$ of the defendant's innocence, and the more tempting is this deviation. These negative observations are formalized by the next proposition (proved in Appendix A.4).

Proposition 2 *There exists no separating PBE in the plea bargain game if:*

i. *either* $f((s|i) \geq 1/w - 1/i)/(1/w - 1/s)$

ii. *or* $f(s|i) \geq s/i$ *and* $f(g|w) < 1 - (g/i)(w/s)$.

Nonetheless, the plea bargain game does have separating PBEs under some parameter values, a set of which is provided by the next proposition. It is obtained through constructing a separating PBE, which I sketch below and detail in Appendix A.6.

Proposition 3 *Separating PBEs exist in the plea bargain game if all the following hold:*

i. $f(s|i) < s/i < 1$,

ii. $f(w|g)/f(w|i) \leq (1/s - 1/g)/(1/s - 1/i)$,

iii. $f(g|s) \leq ((1/g)(w/s) - 1/i)/(1/g - 1/i)$,

iv. $f(g|w) \geq g/i$.

Condition (i) in Proposition 3 rules out the impossibility cases in Proposition 2.[22] With Condition (i), one can also see that $U_d^i(\hat{f}_{d=i}, f_p) \geq U_d^g(\hat{f}_{d=i}, f_p) > 0$ (Ineq. (15), Appendix A.3). That is, if the prosecutor makes an on-path offer and is rejected only by the

[22] Case (i) in Proposition 2 is ruled out because $s/i < 1$ implies, by (5), that $1 < (1/w - 1/i)/(1/w - 1/s)$.

innocent defendant, the innocent defendant gets a positive surplus. And if the guilty defendant pretends to be innocent and rejects the offer then he gets a smaller, positive amount. Then $x_* := U_d^g(\hat{f}_{d=i}, f_p)$ is positive and satisfies the defendant's incentive condition, Ineq. (6).

To satisfy the prosecutor's incentive condition, Ineq. (7), consider separately the two kinds of deviations for the prosecutor, either a downward deviation to offer below x_*, or an upper deviation to offer above x_*.

To deter any downward deviation, let the continuation play be: both types of the defendant reject the deviant offer for sure; in the ensuing criminal trial, the posterior belief about the defendant remains to be the prior f_d by Bayes's rule, while that about the prosecutor is off-path, set to be $\hat{f}_{p=s}$, namely, assigning probability one to her type being strong. With Conditions (i) and (ii) in the proposition, one can show that both types of the prosecutor get zero surplus from this continuation play (Claim A, Appendix A.6). This is the analog of the right-subinterval in Figure 4 with the roles of the two players switched: $d \leftrightarrow p$, $s \leftrightarrow i$ and $w \leftrightarrow g$. For this continuation play to constitute a continuation equilibrium, Conditions (ii) and (iii) together guarantee that both types of the defendant get more from the continuation play than from accepting the prosecutor's deviant offer (Claim B, Appendix A.6).

Any upward deviant offer $x > x_*$ that does not exceed $U_d^i(\hat{f}_{d=i}, f_p)$ in (6) is unprofitable to the prosecutor given the sane continuation play as the one given the on-path offer x_*, because x remains unacceptable to the innocent defendant due to (6), and x yields to the prosecutor less payoff than the on-path x_* when it is accepted by the guilty defendant.

To deter any upward deviation x that exceeds the upper bound $U_d^i(\hat{f}_{d=i}, f_p)$, let the continuation play be: both types of the defendant accept x, and the posterior system remains to be the prior (the posterior about the defendant follows from Bayes's rule, and that of the prosecutor is off-path and hence can be the prior); subsequently, in the off-path event where x is rejected, let the posterior belief about the defendant in the ensuing criminal trial be $\hat{f}_{d=i}$, and that about the prosecutor be the prior f_p. Since $x > U_d^i(\hat{f}_{d=i}, f_p) \geq U_d^g(\hat{f}_{d=i}, f_p)$, accepting x is a best response for both types of the defendant. For the continuation play to be a continuation equilibrium, we need to satisfy (7), which holds if

$$f(g|t_p)(1 - x_*) \geq 1 - x$$

for both types $t_p \in \{s, w\}$ of the prosecutor, as $U_p^{t_p}(\hat{f}_{d=i}, f_p) \geq 0$ for each t_p and both types of the defendant accept x in the continuation play. Since $x > U_d^i(\hat{f}_{d=i}, f_p)$ and $f(g|s) > f(g|w)$

(Ineq. (2), correlation condition), the displayed inequality holds if

$$f(g|w)(1-x_*) \geq 1 - U_d^i(\hat{f}_{d=i}, f_p),$$

which is guaranteed by Condition (iv) in the proposition (Claim C, Appendix A.6). □

The conditions in Proposition 3 are compatible with the strong correlation condition, Ineq. (4). Only Conditions (iii) and (iv) may affect (4), as (i) requires that $f(s|i)$ be small, and (ii) requires that $f(w|g)$ be small and $f(w|i)$ be large, consistent with (4). Plug (iii) and (iv) into the $sf(i|s) < wf(i|w)$ part of (4) to see that (4) is satisfied, by $f(g|s)$ and $f(g|w)$ sufficiently close to their corresponding bounds in (iii) and (iv), if w/s is sufficiently close to one, or g/i sufficiently close to zero (Claim D, Appendix A.6).

Condition (iii) and (iv) restrict the informativeness of the prosecutor's type about the defendant's type: (iii) imposes a cap on $f(g|s)$, which by the correlation between the players should be large, and (iv) keeps a floor on $f(g|w)$, which by the correlation should be small. Without these bounds, when the prosecutor's type is overly informative, the weak prosecutor would assign such a large probability to the innocence of the defendant that she would deviate to a plea deal generous enough for the innocent defendant to accept and plea guilty, thereby upsetting the separating equilibrium.

Since separating PBEs satisfy the normative condition (A), the fact that the plea bargain game admits separating PBEs under some parameter values while the criminal trial without plea bargaining has no BNE that satisfies (A), when the strong correlation condition is satisfied in both models, suggests that adding plea bargaining to criminal trials is a normative improvement, though the improvement is restricted to a large extent due to the difficulty in keeping the innocent from accepting plea deals.

7 Mitigation by the Judge

The restriction highlighted in the previous section can be mitigated by a regulation that need not be overly interfering: Suppose that a judge sets the upper and lower bounds of all the plea deals that the prosecutor is allowed to offer, then the incentive condition for the prosecutor to adhere to a separating PBE would be easier to satisfy, as the upward or downward deviations beyond the bounds are no longer available to her. Then the prospect of having separating PBEs thereby satisfying the normative condition (A) would be improved.

Formally, let us modify the plea bargain game by adding to it an initial stage where a neutral mediator, say a judge, chooses a set $S \subseteq [0,1]$. Then the prosecutor, privately informed of her type, chooses an element from $S \cup \{0\}$ as the plea deal offer, say x. Then the subsequent steps unfold as in the previous model: if $x \neq 0$ then it is offered as the plea deal for the privately informed defendant to choose whether to accept, so that the criminal trial ensues if and only x is rejected; else ($x = 0$) the criminal trial ensues immediately. Given any $S \subseteq [0,1]$, the modified plea bargain game is multistage and so its equilibrium concept is again PBE, and separating PBEs are defined in the same way as in the previous model.

Lemma 3 *If there exists an $S \subseteq [0,1]$ given which the modified plea bargain game has a separating PBE, then when S is replaced by $[U_d^g(\hat{f}_{d=i}, f_p), U_d^i(\hat{f}_{d=i}, f_p)]$, the modified plea bargain game also has a separating PBE.*

Lemma 3 implies that there is no need to consider more heavy-handed interventions than simply setting the upper and lower bounds of plea deals.[23] To prove the lemma, note from the existence of a separating PBE given some S that the defendant's incentive condition, Ineq. (6), is satisfied with some $x_* > 0$ being the on-path plea deal offer. Then (6) implies $U_d^i(\hat{f}_{d=i}, f_p) > 0$. Consequently, from Figures 4 and 5 (or Lemma 7, Appendix A.3) one can see that $f(s|i) < (1/w - 1/i)/(1/w - 1/s)$, which in turn can be shown to imply $U_d^i(\hat{f}_{d=i}, f_p) > U_d^g(\hat{f}_{d=i}, f_p) > 0$ (Lemma 7). Now replace S with $[U_d^g(\hat{f}_{d=i}, f_p), U_d^i(\hat{f}_{d=i}, f_p)]$ and let $x^* := U_d^g(\hat{f}_{d=i}, f_p)$. Note $x^* > 0$. Then construct another separating PBE with x^* being the on-path offer. The defendant's incentive condition, Ineq. (6), is obviously satisfied with x^* replacing x_*. The prosecutor's incentive condition is also easy to satisfy: By the choice of x^* and the new S, if she deviates to some $x \neq x^*$ then either (i) $x^* < x \leq U_d^i(\hat{f}_{d=i}, f_p)$ or (ii) $x = 0$. In Case (i), we have seen, in the reasoning for Proposition 3, that such deviations are unprofitable to the prosecutor. In Case (ii), let the continuation equilibrium be the same as in the case where she chooses $x = 0$ in the original PBE given the original S, so the surplus that the prosecutor of type t_p gets from this deviation is equal to $\tilde{U}_p^{t_p}(0)$.

[23] The most heavy-handed intervention would be for the judge to propose a single plea deal so that the prosecutor either offers it or makes no plea deal offer. While it might appear trivial in light of the positive observation here (the following Proposition 4), such a special case is comparable to the pre-conflict settlement models where the settlement proposal is an unconditional split of the contested good (Kamranzadeh and Zheng [20], Lu, Lu and Riis [22], and Zheng [34]).

Since x_* is the on-path offer in the separating PBE given the original S, it satisfies the prosecutor's incentive condition

$$\tilde{U}_p^{t_p}(0) \leq f(g|t_p)(1-x_*) + f(i|t_p)U_p^{t_p}U_p^{t_p}(\hat{f}_{d=i}, f_p).$$

Since $x^* = U_d^g(\hat{f}_{d=i}, f_p) \leq x_*$, the right-hand side of this inequality does not decrease when x^* replaces x_*. Thus, the prosecutor cannot profit from the deviation in Case (ii) either. □

As explained in the above reasoning, the prosecutor cannot profit from any deviation that is confined to S. The only deviation that can go beyond S is to offer no plea deal (namely, $x = 0$). Thus, the prosecutor's incentive condition, instead of requiring (7) for all deviations $x \in [0,1] \setminus \{x_*\}$, requires (7) only for $x = 0$. That is why the prospect of having separating PBEs, and thereby satisfying the normative condition (A), is much improved by the modification, as observed by our final proposition (proved in Appendix A.7).

Proposition 4 *Separating PBEs exist in the modified plea bargain game if*

i. $f(s|i) < \min\{s/i, (1/w - 1/i)/(1/w - 1/s)\}$ *and*

ii. $f(i|s) \geq (1/g - 1/s)/(1/g - 1/i)$.

Without the condition $f(s|i) < (1/w - 1/i)/(1/w - 1/s)$ required by (i) of the proposition, the innocent would have zero surplus from the criminal trial had he rejected the plea deal in a supposedly separating PBE, and hence he would rather plead guilty. Without the condition $f(s|i) < s/i$ in (i), both types of the prosecutor would get zero surplus from the on-path criminal trial in the said PBE and hence may be tempted to avoid the trial through offering so sweet a deal that can persuade even the innocent to plead guilty. Condition (ii), counterpart to the contrary of $f(s|i) < (1/w - 1/i)/(1/w - 1/s)$ (when the two players switch roles), ensures that each type of the prosecutor gets zero surplus from the off-path event where she offers no plea deal thereby launching the trial without plea bargaining.

It is easy to see that the conditions required by Proposition 4 are strictly weaker than those by its counterpart without the modification, Proposition 3.

8 Conclusion

It is natural to think of plea bargaining as a screening device that helps to separate the guilty from the innocent. In formalizing this idea, the traditional literature on plea bargaining

relies on a reduced form model of the criminal trial, as if it were a nonstrategic final step of the legal system that guarantees higher conviction probabilities for the guilty than for the innocent. Then the screening effect of plea bargaining, in the sense that a guilty defendant is more willing to avoid the trial than his innocent counterpart, becomes a mostly immediate consequence of an assumption that is arguable unrealistic, as the outcome of a trial is often driven by the litigation efforts exerted by the two sides during the trial.

This paper, by contrast, provides a compelling justification for plea bargaining with the proviso that boundaries of plea deals be set beforehand. Replacing the "perfect signal Bayesian court" model of criminal trials in the previous literature by an auction game that captures the all-pay and adversarial nature of litigation, I find that the game admits no equilibrium in which an innocent defendant always commits more litigation resources than his guilty counterpart, and hence the accuracy condition of criminal trials that the previous plea bargaining literature relies on does not hold. This negative observation opens the door for plea bargaining to play an indispensable role in the system to separate the guilty from the innocent. To that end, while unregulated plea bargaining has only a limited role to play due to the restriction needed to keep the prosecutor from offering overly generous deals that pressure the innocent into guilty pleas, the limitation is significantly relaxed when the range of admissible plea deals is regulated by a boundary that can be derived from the parameters.

Such justification for regulated plea bargaining is likely to be valid in a larger set of parameter values than what is reported in this paper. That is because the paper provides only a sufficient condition (in terms of the parameter values) for regulated plea bargaining to satisfy the aforementioned accuracy condition that cannot be satisfied without plea bargaining. To examine the full extent of this contribution brought about by regulated plea bargaining, we need to further the research in two dimensions. First, find a necessary and sufficient condition for the plea bargain game to have separating equilibria, which separate the guilty from the innocent by their responses to the plea deal offers. Second, investigate other kinds of equilibria in the plea bargain game that might also satisfy the accuracy condition.

The first extension direction amounts to finding the most punishing continuation equilibrium in the event where the prosecutor deviates by offering no plea deal at all. The theoretical interest in this problem is that a continuation equilibrium most punishing to one type of the prosecutor need not be punishing enough to deter deviation of the other type, and neither type is a priori harder to be kept from deviation than the other type. That makes

the problem substantially different from the recent conflict-preemption literature such as Lu, Lu and Riis [22] and Zheng [33, 34], where the most punishing continuation equilibrium for the highest type suffices to keep all types from deviation.

The second direction requires investigation about semi-separating equilibria in which there is pooling between the two types of the defendant in their responses to a plea deal offer. Whether such an equilibrium fulfills the normative condition of accuracy would depend on the composition of the types that plead guilty and the continuation play in the criminal trial. That requires full characterization of the bidding equilibrium in the criminal trial given arbitrary posterior systems, including the ones with strong correlation between bidders. In those cases, the nonexistence of monotone equilibria observed in this paper is only a start. More needs to be investigated.

A Proofs

Definition [atoms and gaps] For any cumulative distribution function (cdf) φ on $\mathbb{R}$ and any $b \in \mathbb{R}$, define

$$\varphi(b_-) := \lim_{b' \uparrow b} \varphi(b).$$

Any $b \in \mathbb{R}$ is said to be an *atom* of cdf φ iff φ jumps at b, or equivalently (with cdfs being upper semicontinuous), iff $\varphi(b) - \varphi(b_-) > 0$. For any $b < b'$, (b, b') is said to be a *gap* of φ iff (b, b') is contained in the convex hull of the support of φ and is assigned zero mass by φ, namely, $\varphi(b'_-) - \varphi(b) = 0$.

A.1 A Generalized No-Gap Lemma (Eq. (3))

Given any belief system $(\hat{f}_d, \hat{f}_p)$ where $\hat{f}_d := (\hat{f}(\cdot|s), \hat{f}(\cdot|w))$ represents the belief about the defendant d's type, and $\hat{f}_p := (\hat{f}(\cdot|i), \hat{f}(\cdot|g))$ the belief about the prosecutor p's type, the all-pay auction as a Bayesian game is well-defined, and so is its equilibrium concept Bayesian Nash equilibrium (BNE). Note that $(\hat{f}_d, \hat{f}_p)$ need not be the prior (f_d, f_p). In this subsection I assume, for each player j with distinct types $t_j \neq t'_j$ and for any type t_{-j} of the opponent $-j$,

$$\hat{f}(t_{-j}|t_j) > 0 \iff \hat{f}(t_{-j}|t'_j) > 0. \tag{8}$$

This assumption is satisfied if $\hat{f}$ is derived via Bayes's rule from any nondegenerate prior.

Let $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ denote any BNE of the all-pay auction game given belief system $(\hat{f}_d, \hat{f}_p)$ such that if $\hat{f}(t_{-j}|t_j) = 0$ for some (t_j, t_{-j}) then $\phi_{-j}^{t_{-j}}$ denotes the null function $\varnothing \to \varnothing$ ((8) implies that the player-type $(-j, t_{-j})$ is expected absent by both types of the opponent).

For each player-type (j, t_j), denote the support of the distributional stategy $\phi_j^{t_j}$ by $B_j^{t_j}$, and denote the cdf of the bid submitted by the opponent $-j$ conditional on j's type being t_j by $\Phi_{-j}(\cdot|t_j)$:

$$\Phi_{-j}(b|t_j) := \sum_{t_{-j} \in T_{-j}} \hat{f}(t_{-j}|t_j) \phi_{-j}^{t_{-j}}(b) \tag{9}$$

for all $b \in \mathbb{R}$, where T_{-j} denotes the set of possible types of player $-j$ (e.g., $T_{-d} = T_p = \{s, w\}$). By (8),

$$B_j^{t_j} = \varnothing \iff \left[\forall t_{-j} \in T_{-j} : \hat{f}(t_j|t_{-j}) = 0\right] \iff \left[\exists t_{-j} \in T_{-j} : \hat{f}(t_j|t_{-j}) = 0\right].$$

Thus, the definition (9) implies that the support of $\Phi_{-j}(\cdot|t_j)$ is equal to $\bigcup_{t_{-j} \in T_{-j}} B_{-j}^{t_{-j}}$.

Lemma 4 *For any belief system* $(\hat{f}_d, \hat{f}_p)$ *that satisfies (8) and any BNE* $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ *of the all-pay auction game given* $(\hat{f}_d, \hat{f}_p)$*, there exists* $\overline{b} > 0$ *for which* $\overline{b} = \sup(B_d^g \cup B_d^i) = \sup(B_p^w \cup B_p^s)$.

Proof As explained before the statement of this lemma, the support of $\Phi_d(\cdot|t_p)$ is equal to $B_d^i \cup B_d^g$ for any $t_p \in \{s, w\}$, and the support of $\Phi_p(\cdot|t_d)$ is equal to $B_p^s \cup B_p^w$ for any $t_d \in \{i, g\}$. Then it is trivial that $\sup(B_d^g \cup B_d^i) = \sup(B_p^w \cup B_p^s)$. Thus, denote this common supremum by $\overline{b}$. By the rule of the game, $\overline{b} \geq 0$. We need only to show $\overline{b} > 0$. Suppose, to the contrary, that $\overline{b} = 0$. That is, every bidder-type bids zero for sure at equilibrium, but then obviously each strictly prefers to deviate to bids slightly above zero, contradiction. ■

Lemma 5 *For any BNE* $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ *of the all-pay auction game given any belief system* $(\hat{f}_d, \hat{f}_p)$*, and for any* $j \in \{d, p\}$ *and any* $t_j \in T_j$*, if* (b, b') *is a gap of* $\Phi_{-j}(\cdot|t_j)$ *and* b' *is not an atom of* $\Phi_{-j}(\cdot|t_j)$*, then there exists* $\delta > 0$ *for which* $[b', b' + \delta)$ *is assigned zero mass by* $\phi_j^{t_j}$.

Proof Let $\delta \in (0, b'-b)$ and consider the net gain for (j, t_j) from reducing any bid in $[b', b'+\delta)$ to a bid slightly above b, say $b+\delta/2$. The bidding cost is reduced by $b'-(b+\delta/2) > (b-b')/2 > 0$; and the winning probability for (j, t_j) is reduced by at most

$$\begin{aligned}\Phi_{-j}(b'+\delta|t_j) - \Phi_{-j}(b|t_j) &= \Phi_{-j}(b'+\delta|t_j) - \Phi_{-j}(b'_-|t_j) + \underbrace{\Phi_{-j}(b'_-|t_j) - \Phi_{-j}(b|t_j)}_{=0} \\ &= \Phi_{-j}(b'+\delta|t_j) - \Phi_{-j}(b'_-|t_j) \to 0 \quad \text{as } \delta \to 0,\end{aligned}$$

with the first line due to the assumption that (b, b') is a gap of $\Phi_{-j}(\cdot|t_j)$, and the second line due to the assumption that b' is not an atom of $\Phi_{-j}(\cdot|t_j)$. Thus, the gain for (j, t_j) from this bid reduction is positive for all sufficiently small δ, as claimed. ■

Lemma 6 *For any belief system* $(\hat{f}_d, \hat{f}_p)$ *that satisfies (8) and any BNE* $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ *of the all-pay auction game given* $(\hat{f}_d, \hat{f}_p)$*, there exists* $\overline{b} > 0$ *for which* $[0, \overline{b}] = B_d^g \cup B_d^i = B_p^w \cup B_p^s$*, i.e., (3) holds.*

Proof Suppose, to the contrary, that $B_d^i \cup B_d^g \neq [0, \overline{b}]$. Since $B_d^i \cup B_d^g \subseteq [0, \overline{b}]$ by Lemma 4, $B_d^i \cup B_d^g \subsetneq [0, \overline{b}]$ and $\sup(B_d^i \cup B_d^g) = \overline{b}$. Then by the definition of the support of a cdf, there exist $0 \leq b < b' \leq \overline{b}$ for which $(b, b') \subseteq \mathbb{R} \setminus (B_d^g \cup B_d^i)$. That is, (b, b') is assigned zero mass by both ϕ_d^g and ϕ_d^i. Then (9) implies that (b, b') is a gap of $\Phi_d(\cdot|t_p)$ for each $t_p \in \{s, w\}$. Choose b' to be maximal, so that there exists no $b'' \in (b', \overline{b})$ for which $(b, b'') \subseteq \mathbb{R} \setminus (B_d^g \cup B_d^i)$.

As distribution functions are upper semicontinuous, $b' \in B_d^{t_d}$ for some $t_d \in \{g, i\}$. That is, there exists $t_d \in \{g, i\}$ for which

$$\forall \epsilon > 0 : \ \phi_d^{t_d}(b' + \epsilon) - \phi_d^{t_d}(b'_-) > 0, \tag{10}$$

For each $t_p \in \{s, w\}$, since (b, b') is a gap of $\Phi_d(\cdot|t_p)$, no element of (b, b') belongs to $B_p^{t_p}$, as none is a best response for player-type (p, t_p). Thus, (b, b') is assigned zero mass by both ϕ_p^s and ϕ_p^w. Then (9) implies that (b, b') is a gap of both $\Phi_p(\cdot|i)$ and $\Phi_p(\cdot|g)$. Consequently, Lemma 5 implies that for each $t_d \in \{g, i\}$, either b' is an atom of $\Phi_p(\cdot|t_d)$, or (10) does not hold. Since (10) does hold for some $t_d \in \{g, i\}$, as proved before, it follows that b' is an atom of $\Phi_p(\cdot|t_d)$ for some $t_d \in \{g, i\}$. Consequently, by (9),

$$b' \text{ is an atom of } \phi_p^{t_p} \text{ for some } t_p \in \{s, w\}. \tag{11}$$

Since both ϕ_p^s and ϕ_p^w are constant on (b, b') ((b, b') assigned zero mass by both ϕ_p^s and ϕ_p^w), the jump of $\phi_p^{t_p}$ at b' cannot be smoothed away by the other $\phi_p^{t'_p}$ ($t'_p \in T_p \setminus \{t_p\}$) when the two are combined to have Φ_p via (9). Thus, b' is an atom of $\Phi_p(\cdot|t'_d)$ for each $t'_d \in \{i, g\}$. Consequently, for each $t'_d \in \{i, g\}$, b' is not a best response for (d, t'_d), and hence not an atom of $\phi_d^{t'_d}$. Apply (9) again to see that b' is not an atom of $\Phi_d(\cdot|t_p)$. This, combined with Lemma 5 and the fact that (b, b') is a gap of $\Phi_d(\cdot|t_p)$, implies that $b' \notin B_p^{t_p}$. But that contradicts (11).

Thus, $B_d^g \cup B_d^i = [0, \overline{b}]$ is proved by contradiction. The case for $B_p^w \cup B_p^s$ is the same. ■

A.2 Proposition 1

Let $(\phi_d^i, \phi_d^g, \phi_p^s, \phi_p^w)$ denote any BNE of the auction so that $\phi_j^{t_j}$ denotes the cdf according to which player-type (j, t_j) (player j of type t_j) chooses a bid to submit, and denote the support of $\phi_j^{t_j}$ by $B_j^{t_j}$. By Lemma 6, there exists a $\overline{b} > 0$ associated to the BNE for which

$$B_d^i \cup B_d^g = B_p^s \cup B_p^w = \left[0, \overline{b}\right]. \tag{12}$$

For each player-type (j, t_j), denote the cdf of the bid submitted by the opponent $-j$ conditional on j's type being t_j by $\Phi_{-j}(\cdot|t_j)$ as in (9), with the $\hat{f}$ there being the prior f here.

Suppose, to the contrary of the claim, that the defendant's equilibrium strategy (ϕ_d^i, ϕ_d^g) is monotone. Then, by (12), B_d^i and B_d^g are each an interval, and $\sup B_d^g \leq \inf B_d^i$. It follows that B_d^i is not singleton; otherwise, $B_d^i = \{\overline{b}\}$ is an atom (mass point) of $\Phi_d(\cdot|t_p)$ for each

$t_p \in \{s, w\}$ ($f(i|t_p) > 0$ by nondegeneracy of f) and then player-type (p, t_p) would replace bids slightly below $\overline{b}$ by bids slightly above $\overline{b}$, contradicting (12). Thus, $B_d^i = [b_*, \overline{b}]$ for some $0 \leq b_* < \overline{b}$. By (12), $\left(b_*, \overline{b}\right) \subset B_p^s \cup B_p^w$. Now that $(b_*, \overline{b}] \subseteq B_d^i \setminus B_d^g$,

$$\Phi_d(b|t_p) = f(g|t_p) + f(i|t_p)\phi_d^i(b)$$

for each $b \in (b_*, \overline{b}]$ and each $t_p \in \{s, w\}$.

Claim 1: ϕ_d^i has no atom on $(b_*, \overline{b}]$. Otherwise, $\Phi_d(\cdot|t_p)$ ($= f(g|t_p) + f(i|t_p)\phi_d^i(\cdot)$ on $(b_*, \overline{b}]$ as noted before) has an atom in $(b_*, \overline{b}]$ for each $t_p \in \{s, w\}$. But then both types s and w of player p would prefer bids slightly above the atom to bids slightly below the atom, and hence $B_p^s \cup B_p^w$ is not an interval $[0, \overline{b}]$, contradicting (12).

Claim 2: If $b, b' \in (b_*, \overline{b}]$, $b' \in B_p^s$ and $b \in B_p^w$, then $b' \leq b$. By the monotonicity supposition, $\Phi_d(b'|s) = f(g|s) + f(i|s)\phi_d^i(b')$ and $\Phi_d(b|w) = f(g|w) + f(g|w)\phi_d^i(b)$. By Claim 1, $\Phi_d(b'|s)$ is equal to the probability for (p, s) to win with bid b', and $\Phi_d(b|w)$, the probability for (p, w) to win with bid b. Thus, the rationalizability of b' for (p, s), and that of b for (p, w), require that

$$f(g|s) + f(i|s)\phi_d^i(b') - \frac{b'}{s} \geq f(g|s) + f(i|s)\phi_d^i(b) - \frac{b}{s} \tag{13}$$

$$f(g|w) + f(i|w)\phi_d^i(b) - \frac{b}{w} \geq f(g|w) + f(i|w)\phi_d^i(b') - \frac{b'}{w}. \tag{14}$$

The two inequalities are equivalent to

$$\begin{aligned} sf(i|s)\phi_d^i(b') - b' &\geq sf(i|s)\phi_d^i(b) - b \\ wf(i|w)\phi_d^i(b) - b &\geq wf(i|w)\phi_d^i(b') - b'. \end{aligned}$$

Sum them to obtain

$$\left(\phi_d^i(b') - \phi_d^i(b)\right)\left(sf(i|s) - wf(i|w)\right) \geq 0.$$

Thus, $\phi_d^i(b') \leq \phi_d^i(b)$ because $sf(i|s) - wf(i|w) < 0$ by (4) (strong correlation). Note that ϕ_d^i is strictly increasing on $(b_*, \overline{b}]$ (due to (12) and the monotonicity supposition $(b_*, \overline{b}] \subset B_d^i \setminus B_d^g$). It then follows that $b' \leq b$, as claimed.

Claim 3: The set of $(b_*, \overline{b}] \cap B_p^s \cap B_p^w$ cannot contain more than one point. Suppose, to the contrary, that $b_* < b < b' \leq \overline{b}$ and $b, b' \in (b_*, \overline{b}] \cap B_p^s \cap B_p^w$. Note from Claim 1 that the probability for any type t_p of player p to win with any bid $b'' \in (b_*, \overline{b}]$ is equal to $\Phi_d(b''|t_p)$. Thus, to rationalize any two distinct bids b and b' in $(b_*, \overline{b}]$ for both types s and w of player p,

$$\frac{1}{sf(i|s)} = \frac{\phi_d^i(b') - \phi_d^i(b)}{b' - b} = \frac{1}{wf(i|w)}.$$

which is impossible due to the strong correlation inequality (4).

Claim 4: $(b_*, \bar{b}] \subset B_p^w \setminus B_p^s$. Suppose not, then by Claims 2 and 3, there exists $b^* \in (b_*, \bar{b}]$ for which $(b_*, b^*) \subset B_p^s \setminus B_p^w$. Note that $\Phi_p(\cdot|i)$ has no atom on $(b_*, \bar{b}]$, otherwise (12) would be contradicted. It follows that $\Phi_p(b|i)$ is equal to the probability for (d, i) to win with any bid $b \in (b_*, \bar{b}]$. Thus, to any $b \in (b_*, b^*)$ for (d, i) (the type i of player d), we need

$$\dot{\Phi}_p(b|i) = f(s|i)\dot{\phi}_p^s(b) - \frac{1}{i} = 0, \quad \text{namely,} \quad \dot{\phi}_p^s(b) = \frac{1}{if(s|i)}.$$

Thus, in (b_*, b^*) there is no atom for ϕ_p^s and hence no atom for $\Phi_p(\cdot|t_d)$ for each $t_d \in \{g, i\}$. Now consider player-type (d, g), who is supposed to submit no bid above b_* according to the monotonicity supposition. The supposition requires that b_* is not an atom of $\Phi_p(\cdot|g)$, otherwise (d, g) would replace the bids equal to or slightly below b_* by bids slightly above b_*. Thus, $\Phi_p(\cdot|g)$ is continuous at b_*, and for every $b \in (b_*, b^*)$,

$$\dot{\Phi}_p(b|g) = f(s|g)\dot{\phi}_p^s(b) = \frac{f(s|g)}{if(s|i)} > \frac{1}{g},$$

with the inequality due to (4) (strong correlation). Since $\Phi_p(\cdot|g)$ has no atom in $[b_*, b^*)$, it is equal to the winning probability for (d, g) with bids in this interval. Thus, the inequality displayed above implies that (d, g) strictly prefers bids in $(b_*, b^*]$ to the bid b^*, contradicting the monotonicity supposition. This contradiction establishes the claim.

Claim 5: $B_p^s \subseteq [0, b_*]$ and $0 < b_*$. Claim 4 implies $B_p^s \subseteq [0, b_*]$ directly. If $0 = b_*$ then $\phi_p^s(0) = \phi_d^g(0) = 1$ and hence $\Phi_p(0|t_d) \geq f(s|t_d) > 0$ and $\Phi_d(0|t_p) = f(g|t_p) > 0$ for all t_d and all t_p by the nondegeneracy assumption of f. It follows that each player-type (j, t_j) faces an atom at the zero bid from the opponent, and hence would replace the zero bid that he is supposed to submit with positive probability by bids slightly above zero: contradiction.

Claim 6: If $b, b' \in (0, b_*]$, $b' \in B_p^s$ and $b \in B_p^w$, then $b' \geq b$. Mimic the proof of Claim 2, with the role of i there played by g here, to obtain

$$\left(\phi_d^g(b') - \phi_d^g(b)\right)\left(sf(g|s) - wf(g|w)\right) \geq 0.$$

Since $sf(g|s) > sf(g|w) > wf(g|w)$ by (2) (the correlation condition), the inequality displayed above implies $\phi_d^g(b') \geq \phi_d^g(b)$ and hence (as in Claim 2) $b' \geq b$.

Claim 7: $(0, b_*] \subset B_p^w \setminus B_p^s$. Suppose not, then by Claim 6 there exists $b' \in (0, b_*]$ for which $(b', b_*] \subset B_p^s \setminus B_p^w$. Then mimic the proof of Claim 4 to obtain

$$\dot{\phi}_d^g(b) = \frac{1}{sf(g|s)}$$

for all $b \in (b', b_*)$. Then consider player p of type w, who, as established previously, submits bids in $(b_*, \bar{b}]$ but not any bid in (b', b_*). If (p, w) reduces her bid from b_* to any bid in (b', b_*), the expected payoff for (p, w) increases at the rate

$$-\left(\dot{\Phi}_d(b|w) - \frac{1}{w}\right) = -\left(f(g|w)\dot{\phi}_d^g(b) - \frac{1}{w}\right) = -\left(\frac{f(g|w)}{sf(g|s)} - \frac{1}{w}\right) > -\left(\frac{1}{s} - \frac{1}{w}\right) > 0,$$

with the first inequality due to (2) (correlation). Thus, (p, w) would deviate to the bids in (b', b_*), contradiction.

By Claim 7, $B_p^s = \{0\}$ and hence (p, s) gets zero surplus. But then (p, s) would deviate to bidding in $(0, b_*)$ thereby getting a positive expected payoff: Since bids in $(0, b_*)$ are submitted only by (p, w) and (d, g) at equilibrium, to rationalize any such bid b for (p, w),

$$\dot{\phi}_d^g(b) = \frac{1}{wf(g|w)}.$$

Consequently, by increasing the bid from zero to the interval $(0, b_*)$, player p of type s increases her expected payoff at the rate

$$\frac{f(g|s)}{wf(g|w)} - \frac{1}{s} > \frac{1}{w} - \frac{1}{s} > 0,$$

with the first inequality due to (2) (correlation). Thus, $B_p^s \neq \{0\}$, contradiction. Consequently, the monotonicity supposition cannot hold, as asserted. ■

A.3 Figures 4 and 5

Lemma 7 *For any posterior belief $\hat{f}_p := (\hat{f}(\cdot|i), \hat{f}(\cdot|g))$ of the prosecutor's type, the all-pay auction given belief system $(\hat{f}_{d=i}, \hat{f}_p)$ has a unique BNE and:*

i. if $\hat{f}(s|i) \geq s/i$ then

$$\begin{aligned} U_p^s(\hat{f}_{d=i}, \hat{f}_p) &= U_p^w(\hat{f}_{d=i}, \hat{f}_p) = 0 \\ U_d^i(\hat{f}_{d=i}, \hat{f}_p) &= 1 - s/i \\ U_d^g(\hat{f}_{d=i}, f_p) &= 1 - s/g < U_d^i \hat{f}_{d=i}, f_p) \quad \text{if } \hat{f}_p = f_p; \end{aligned}$$

ii. if $\hat{f}(s|i) \geq \frac{1/w - 1/i}{1/w - 1/s}$ then

$$\begin{aligned} U_d^i(\hat{f}_{d=i}, \hat{f}_p) &= 0 \\ U_d^g(\hat{f}_{d=i}, f_p) &= 0 \quad \text{if } \hat{f}_p = f_p \\ U_p^s(\hat{f}_{d=i}, \hat{f}_p) &= 1 - i/s \\ U_p^w(\hat{f}_{d=i}, \hat{f}_p) &= 1 - \frac{i\hat{f}(s|i)}{s} - \frac{i\hat{f}(w|i)}{w}; \end{aligned}$$

iii. if $\hat{f}(s|i) < \min\left\{\frac{s}{i}, \frac{1/w-1/i}{1/w-1/s}\right\}$ *then*

$$\begin{aligned} U_p^w(\hat{f}_{d=i}, \hat{f}_p) &= 0 \\ U_d^i(\hat{f}_{d=i}, \hat{f}_p) &= \hat{f}(w|i) - \frac{w}{i}\left(1 - \frac{i\hat{f}(s|i)}{s}\right) \end{aligned}$$

and, if $\hat{f}_p = f_p$,

$$\begin{aligned} U_d^g(\hat{f}_{d=i}, f_p) &= \max\left\{ \begin{array}{l} f(w|g)\left(1 - \frac{1}{if(w|i)} w\left(1 - \frac{if(s|i)}{s}\right)\right), \\ 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) \end{array} \right\} \\ U_d^i(\hat{f}_{d=i}, f_p) &> U_d^g(\hat{f}_{d=i}, f_p) > 0. \end{aligned} \tag{15}$$

Note that the three cases listed in Lemma 7 correspond to Figures 4 and 5. By (5) noted before, Case (i) in the lemma implies $1 \geq \hat{f}(s|i) \geq s/i$ and hence corresponds to the right interval in Figure 5. By the same token, Case (ii) implies $1 \geq (1/w - 1/i)/(1/w - 1/s)$ and hence $s/i \geq 1$, so it corresponds to the right interval in Figure 4. In Case (iii), $\min\left\{\frac{s}{i}, \frac{1/w-1/i}{1/w-1/s}\right\} = s/i$ if and only if $s/i \leq 1$; hence it is the left interval in either figure.

Proof of the lemma Let $(\phi_d^i, \phi_p^s, \phi_p^w)$ denote any BNE of the all-pay auction given belief system $(\hat{f}_{d=i}, \hat{f}_p)$. Define $\Phi_{-j}(\cdot|t_j)$ as in (9) except that the prior f there is replaced by the posterior $\hat{f}$ here. Since $\hat{f}_{d=i}(g|t_p) = 0$ for each $t_p \in \{s, w\}$), the support B_d^g of the strategy for (d, g) is empty. By Lemma 4, $B_d^i = B_p^s \cup B_p^w = [0, \overline{b}]$ for some $\overline{b} > 0$.

Observe that the prosecutor's strategy (ϕ_p^s, ϕ_p^w) is monotone on $[0, \overline{b}]$. To see that, note $\Phi_d(\cdot|t_p) = \phi_d^i$ on $[0, \overline{b}]$ for each $t_p \in \{s, w\}$ and hence $(0, \overline{b}]$ contains neither atom nor gap of $\Phi_d(\cdot|t_p)$ (otherwise each type of the prosecutor would deviate and hence $B_p^s \cup B_p^w \neq [0, \overline{b}]$. It follows that $\Phi_d(b|t_p)$ is equal to the probability with which (p, t_p) wins from submitting any bid in $[0, \overline{b}]$. Then mimic the reasoning starting from (13) and (14), with the $f(g|s) + f(i|s)\phi_d^i(b')$ there replaced by $\phi_d^i(b')$, $f(g|s) + f(i|s)\phi_d^i(b)$ replaced by $\phi_d^i(b)$, and likewise for $f(g|w) + f(i|w)\phi_d^i(b)$ and $f(g|w) + f(i|w)\phi_d^i(b')$. That gives us, for any $b' \in B_p^s$ and any $b \in B_p^w$,

$$\left(\phi_d^i(b') - \phi_d^i(b)\right)(s - w) \geq 0$$

and hence $\phi_d^i(b') \geq \phi_d^i(b)$. Since ϕ_d^i has no gap, we have $b' \geq b$.

It follows that there is some $0 \leq b_* < \overline{b}$ for which $B_p^w = [0, b_*]$ and $B_p^s = [b_*, \overline{b}]$. We shall solve for b_* and $\overline{b}$.

Since any bid $b \in (b_*, \overline{b})$ is submitted only by (p,s) and (d,i), the expected payoff is equal to $\phi_d^i(b) - b/s$ for (p,s), and equal to $\hat{f}(w|i) + \phi_p^s(b)\hat{f}(s|i) - b/i$ for (d,i), in submitting any $b \in (b_*, \overline{b})$. For these b to be the mutual best responses between (p,s) and (d,i), their expected payoffs should each be equal to a constant on $(b_*, \overline{b})$. Thus,

$$\begin{aligned} \dot{\phi}_d^i(b) &= \frac{1}{s} \\ \dot{\phi}_p^s(b) &= \frac{1}{i\hat{f}(s|i)} \end{aligned}$$

for all $b \in (b_*, \overline{b})$.

Analogously, any bid $b \in (0, b_*)$ is submitted only by (p,w) and (d,i), and hence the best response condition requires, for all $b \in (0, b_*)$, that

$$\begin{aligned} \dot{\phi}_d^i(b) &= \frac{1}{w}, \\ \dot{\phi}_p^w(b) &= \frac{1}{i\hat{f}(w|i)}. \end{aligned}$$

For (p,s) to not deviate to bids below b_*, it suffices to have $\dot{\phi}_d^i(b) - 1/s \geq 0$ for all $b \in (0, b_*)$. By the equations displayed above, $\dot{\phi}_d^i(b) = 1/w > 1/s$ for all such b, hence (p,s) cannot profit from the deviation. Likewise, since $\dot{\phi}_d^i(b) = 1/s < 1/w$ for all $b \in (b_*, \overline{b})$, neither does (p,w) want to deviate to bids above b_*.

To pin down $(\phi_d^i, \phi_p^s, \phi_p^w)$, note from the fact $B_p^s = [b_*, \overline{b}] \subseteq [0, \overline{b}] = B_d^i$ that

$$\frac{1}{i\hat{f}(s|i)} < \frac{1}{s} \iff \dot{\phi}_p^s < \dot{\phi}_d^i \iff \phi_p^s(b_*) = 0 \iff b_* > 0.$$

Thus, let us bifurcate.

Case 1: $i\hat{f}(s|i) \geq s$ Then $b_* = 0$ and the weak prosecutor (p,w) bids zero for sure. With the prosecutor bidding zero with a positive probability, the defendant, in best responding, bids zero with zero probability, namely, $\phi_d^i(0) = \phi_d^i(b_*) = 0$. Thus,

$$1 = \phi_d^i(\overline{b}) - \phi_d^i(b_*) = (\overline{b} - b_*)\dot{\phi}_d^i|_{(b_*, \overline{b})} = \overline{b}/s.$$

Hence $\overline{b} = s$. Since $1 - \phi_p^s(0) = 1 - \phi_p^s(b_*) = (\overline{b} - b_*)\dot{\phi}_p^s|_{(b_*, \overline{b})} = \frac{s-0}{i\hat{f}(s|i)}$,

$$\phi_p^s(0) = 1 - \frac{s}{i\hat{f}(s|i)}.$$

This pins down the equilibrium. Thus the surpluses in this equilibrium are:

$$\begin{aligned} U_p^s(\hat{f}_{d=i}, \hat{f}_p) &= U_p^w(\hat{f}_{d=i}, \hat{f}_p) = 0 \\ U_d^i(\hat{f}_{d=i}, \hat{f}_p) &= \hat{f}(w|i) + \hat{f}(s|i)\phi_p^s(0) = \hat{f}(w|i) + \hat{f}(s|i)\left(1 - \frac{s}{i\hat{f}(s|i)}\right) = 1 - \frac{s}{i}. \end{aligned}$$

While assumed to be of zero probability by the posterior $\hat{f}_{d=i}$, the guilty defendant (d, g), expecting (ϕ_p^s, ϕ_p^w), gets the expected payoff

$$\begin{aligned} U_d^g(\hat{f}_{d=i}, \hat{f}_p) &= \max\left\{\hat{f}(w|g) + \hat{f}(s|g)\phi_p^s(0), 1 - \bar{b}/g\right\} \\ &= \max\left\{\hat{f}(w|g) + \hat{f}(s|g)\left(1 - \frac{s}{i\hat{f}(s|i)}\right), 1 - \frac{s}{g}\right\} \\ &= \max\left\{1 - \frac{s\hat{f}(s|g)}{i\hat{f}(s|i)}, 1 - \frac{s}{g}\right\} \\ &\overset{(4)}{=} 1 - s/g \quad \text{if } \hat{f} = f. \end{aligned}$$

(Note from (4) that $s \leq if(s|i) < g(f(s|g) < g.)$ Thus, Part (i) of Lemma 7 is proved.

Case 2: $i\hat{f}(s|i) < s$ That is, $b_* > 0$. Note from $B_p^s = [b_*, \bar{b}]$ that $1 = (\bar{b} - b_*)\dot{\phi}_p^s(b)$ for any $b \in (b_*, \bar{b})$. By the $dt\phi_p^s$ calculated previously, that means $1 = (\bar{b} - b_*)/(i\hat{f}(s|i))$, namely,

$$\bar{b} - b_* = i\hat{f}(s|i).$$

Plug this into the fact $1 - \phi_d^i(b_*) = (\bar{b} - b_*)\dot{\phi}_d^i(b)|_{b>b_*} = (\bar{b} - b_*)/s$ to obtain

$$\phi_d^i(b_*) = 1 - \frac{i\hat{f}(s|i)}{s}.$$

To figure out whose bid distribution has an atom at zero, let b_j be the infimum of the support of bidder j's bid distribution if the nonnegativity constraint for bids were relaxed. That is,

$$1 = \phi_p^w(b_*) - \phi_p^w(b_p) = (b_* - b_p)\dot{\phi}_p^w(b)|_{b<b_*} = \frac{b_* - b_p}{i\hat{f}(w|i)} \quad \Rightarrow \quad b_p = b_* - i\hat{f}(w|i)$$

$$\phi_d^i(b_*) - \phi_d^i(b_d) = 1 - \frac{i\hat{f}(s|i)}{s} = (b_* - b_d)\dot{\phi}_d^i(b)|_{b<b_*} = \frac{b_* - b_d}{w} \quad \Rightarrow \quad b_d = b_* - w\left(1 - \frac{i\hat{f}(s|i)}{s}\right).$$

Subcase 2a: $b_p \geq b_d$ Namely, $i\hat{f}(w|i) \leq w\left(1 - \frac{i\hat{f}(s|i)}{s}\right)$. Then

$$\phi_p^w(0) = 0 \leq \phi_d^i(0).$$

Consequently, $1 = \phi_p^w(b_*) - \phi_p^w(0) = b_*/(i\hat{f}(w|i))$ and hence

$$\begin{aligned} b_* &= i\hat{f}(w|i) \\ \overline{b} &= i\hat{f}(w|i) + i\hat{f}(s|i) = i. \end{aligned}$$

Then $\phi_d^i(b_*) - \phi_d^i(0) = 1 - \frac{i\hat{f}(s|i)}{s} - \phi_d^i(0) = b_*/w = i\hat{f}(w|i)/w$ and hence

$$\phi_d^i(0) = 1 - \frac{i\hat{f}(s|i)}{s} - \frac{i\hat{f}(w|i)}{w}.$$

Denote the surplus for bidder-type (j, t_j) at this equilibrium by $U_j^{t_j}(\hat{f}_{d=i}, \hat{f}_p)$. Then

$$\begin{aligned} U_p^w(\hat{f}_{d=i}, \hat{f}_p) &= \phi_d^i(0) = 1 - \frac{i\hat{f}(s|i)}{s} - \frac{i\hat{f}(w|i)}{w} \\ U_p^s(\hat{f}_{d=i}, \hat{f}_p) &= 1 - \frac{\overline{b}}{s} = 1 - \frac{i}{s} > 1 - \frac{i}{s}\left(\hat{f}(s|i) + \frac{s}{w}\hat{f}(w|i)\right) = U_p^w(\hat{f}_{d=i}, \hat{f}_p) \\ U_d^i(\hat{f}_{d=i}, \hat{f}_p) &= 1 - \frac{\overline{b}}{i} = 1 - i/i = 0. \end{aligned}$$

To calculate $U_d^g(\hat{f}_{d=i}, \hat{f}_p)$, note that the expected payoff for (d, g) given this equilibrium is equal to $\hat{f}(w|g)\phi_p^w(b) - b/g$ for all $b \in (0, b_*)$, and $\hat{f}(w|g) + \hat{f}(s|g)\phi_p^s(b) - b/g$ for all $b \in (b_*, \overline{b}]$. Thus, the derivative of the expected payoff with respect to b is equal to

$$\hat{f}(w|g)\dot{\phi}_p^w(b)|_{b<b_*} - \frac{1}{g} = \frac{\hat{f}(w|g)}{i\hat{f}(w|i)} - \frac{1}{g}$$

for all $b \in (0, b_*)$, and

$$\hat{f}(s|g)\dot{\phi}_p^s(b)|_{b>b_*} - \frac{1}{g} = \frac{\hat{f}(s|g)}{i\hat{f}(s|i)} - \frac{1}{g}$$

for all $b \in (b_*, \overline{b})$. If $\hat{f}(\cdot|t_d) = f(\cdot|t_d)$ then one can prove from (4) and $i > g$ that

$$gf(w|g) < if(w|i). \tag{16}$$

Then the derivative on the first line is negative, and by (4) the derivative on the second line is positive. Thus,

$$U_d^g(\hat{f}_{d=i}, f_p) = \max\left\{\lim_{b\downarrow 0}\left(f(w|g)\phi_p^w(b) - b/g\right), 1 - \overline{b}/g\right\} = \max\{f(w|g)\phi_p^w(0), 1 - i/g\} = 0.$$

Since $i\hat{f}(w|i) \leq w\left(1 - \frac{i\hat{f}(s|i)}{s}\right) \iff \hat{f}(s|i) \geq (1/w - 1/i)/(1/w - 1/s)$, Part (ii) of Lemma 7 is proved.

Subcase 2b: $b_p \le b_d$ Namely, $i\hat{f}(w|i) \ge w\left(1 - \frac{i\hat{f}(s|i)}{s}\right)$. Then

$$\phi_p^w(0) \ge 0 = \phi_d^i(0).$$

Consequently, $b_d = 0$, i.e., $b_* - w(1 - i\hat{f}(s|i)/s) = 0$ and hence

$$\begin{aligned} b_* &= w\left(1 - \frac{i\hat{f}(s|i)}{s}\right) \\ \overline{b} &= w\left(1 - \frac{i\hat{f}(s|i)}{s}\right) + i\hat{f}(s|i). \end{aligned}$$

Then

$$\phi_p^w(b_*) - \phi_p^w(0) = 1 - \phi_p^w(0) = \frac{b_*}{i\hat{f}(w|i)} = \frac{1}{i\hat{f}(w|i)} w\left(1 - \frac{i\hat{f}(s|i)}{s}\right)$$

and hence

$$\phi_p^w(0) = 1 - \frac{1}{i\hat{f}(w|i)} w\left(1 - \frac{i\hat{f}(s|i)}{s}\right).$$

Then

$$\begin{aligned} U_p^w(\hat{f}_{d=i}, \hat{f}_p) &= 0 \\ U_p^s(\hat{f}_{d=i}, \hat{f}_p) &= \phi_d^i(b_*) - \frac{b_*}{s} = (b_* - 0)\dot{\phi}_d^i(b)|_{b<b_*} - \frac{b_*}{s} = \frac{b_*}{w} - \frac{b_*}{s} = 1 - \frac{i\hat{f}(s|i)}{s} - \frac{w}{s}\left(1 - \frac{i\hat{f}(s|i)}{s}\right) \\ &= 1 - \frac{w}{s} - \frac{i\hat{f}(s|i)}{s}\left(1 - \frac{w}{s}\right) = \left(1 - \frac{w}{s}\right)\left(1 - \frac{i\hat{f}(s|i)}{s}\right) > 0 \qquad (17) \\ U_d^i(\hat{f}_{d=i}, \hat{f}_p) &= \lim_{b\downarrow 0}\left(\hat{f}(w|i)\phi_p^w(b) - b/i\right) = \hat{f}(w|i)\phi_p^w(0) = \hat{f}(w|i) - \frac{w}{i}\left(1 - \frac{i\hat{f}(s|i)}{s}\right) \ge 0. \end{aligned}$$

To calculate $U_d^g(\hat{f}_{d=i}, \hat{f}_p)$, note from the fact that the derivative of the opponent's bid distribution is constant on $(0, b_*)$ and on $(b_*, \overline{b})$ that

$$\begin{aligned} U_d^g(\hat{f}_{d=i}, \hat{f}_p) &= \max\left\{\lim_{b\downarrow 0}\left(\hat{f}(w|g)\phi_p^w(b) - b/g\right), 1 - \overline{b}/g, \hat{f}(w|g) - b_*/g\right\} \\ &= \max\left\{\hat{f}(w|g)\phi_p^w(0), 1 - \frac{\overline{b}}{g}, \hat{f}(w|g) - \frac{w}{g}\left(1 - \frac{i\hat{f}(s|i)}{s}\right)\right\} \\ &= \max\left\{f(w|g)\phi_p^w(0), 1 - \frac{\overline{b}}{g}\right\} \quad \text{if } \hat{f} = f \text{ and } 1 - \tfrac{if(s|i)}{s} \ge 0, \qquad (18) \end{aligned}$$

with the third line due to

$$f(w|g)\phi_p^w(0) = f(w|g) - \frac{f(w|g)}{if(w|i)} w \underbrace{\left(1 - \frac{if(s|i)}{s}\right)}_{>0\text{: current Case 2}} \overset{(16)}{>} f(w|g) - \frac{1}{g} w\left(1 - \frac{if(s|i)}{s}\right).$$

Note $U_d^g(\hat{f}_{d=i}, \hat{f}_p) \geq 0$ because $\phi_p^w(0) \geq 0$, and strictly so if the inequality that defines this case (Subcase 2b) is strict. Among the items inside the $\max\{\cdots\}$, $1 - (\bar{b}/i)(i/g) < 1 - \bar{b}/i = U_d^i(\hat{f}_{d=i}, \hat{f}_p)$ since $i > g$ and $[0, \bar{b}] = \operatorname{supp} \phi_d^i$; if $\hat{f}(\cdot|t_d) = f(\cdot|t_d)$, then

$$\begin{aligned} f(w|g)\phi_p^w(0) &= f(w|g)\left(1 - \frac{1}{if(w|i)} w \left(1 - \frac{if(s|i)}{s}\right)\right) \\ &\overset{(2)}{<} f(w|i)\left(1 - \frac{1}{if(w|i)} w \left(1 - \frac{if(s|i)}{s}\right)\right) = U_d^i(\hat{f}_{d=i}, f_p). \end{aligned}$$

Since $i\hat{f}(w|i) > w\left(1 - \frac{i\hat{f}(s|i)}{s}\right) \iff \hat{f}(s|i) < (1/w - 1/i)/(1/w - 1/s)$ and we are currently outside Case 1 (not $i\hat{f}(s|i) \leq s$), the current case corresponds to Part (iii) of Lemma 7. Plug the solution for $\bar{b}$ obtained above into (18) and the proof is complete. ■

A.4 Proposition 2

The explanation preceding the statement of this proposition has shown that Condition (i) in the proposition implies impossibility of separating PBE. To prove that Condition (ii) also implies such, the next lemma suffices.

Lemma 8 *If $f(s|i) \geq s/i$ then in any separating PBE the incentive condition for the prosecutor in any separating PBE (Ineq. (7)) requires for each $t_p \in \{s, w\}$ that*

$$f(g|t_p) \geq \frac{g}{i} \cdot \frac{w}{s}.$$

Proof By $f(s|i) \geq s/i$, Part (i) of Lemma 7 applies. Thus, in the on-path event where the equilibrium offer x_* is rejected, both types of the prosecutor get zero surplus, and the surpluses for the defendant are $U_d^i(\hat{f}_{d=i}, f_p) = 1 - s/i$ and $U_d^g(\hat{f}_{d=i}, f_p) = 1 - s/g$. Then (7) requires, for any deviation $x \neq x_*$ and each $t_p \in \{s, w\}$, that

$$f(g|t_p)(1 - x_*) + f(i|t_p) \underbrace{U_p^{t_p}(\hat{f}_{d=i}, f_p)}_{=0} \geq \tilde{U}_p^{t_p}(x).$$

This, combined with $x_* \geq U_d^g(f, \overline{\mu}_d)$ (due to the incentive condition (6) for separating PBEs) and $U_d^g(f, \overline{\mu}_d) = 1 - s/g$, implies for each $t_p \in \{s, w\}$ that

$$f(g|t_p) \cdot \frac{s}{g} \geq \tilde{U}_p^{t_p}(x). \tag{19}$$

Let $x > 1 - w/i$. Let $e(x) := (\sigma_d^{i,x}, \sigma_d^{g,x}, \hat{f}_p^x, \hat{f}_d^x, \phi_p^{s,x}, \phi_p^{w,x}, \phi_d^{i,x}, \phi_d^{g,x})$ be the continuation equilibrium (cf. Section 5 for the noation), as part of the separating PBE under consideration, in the event that the prosecutor deviates by offering x instead of x_*.

Claim: In $e(x)$ both types of the defendant accept the offer x for sure. Otherwise, $\sigma_d^{t_d,x} > 0$ for at least one $t_d \in \{i, g\}$. That is, on the path of $e(x)$, at least one type t_d of the defendant rejects the offer x with a positive probability. By Bayes's rule and the assumption that the prior f is nondegenerate,

$$\left[\exists t_p \in \{s, w\} : \hat{f}_d^x(t_d|t_p) > 0\right] \Rightarrow \sigma_d^{t_d,x} > 0 \Rightarrow \left[\forall t_p \in \{s, w\} : \hat{f}_d^x(t_d|t_p) > 0\right].$$

Since $\sigma_d^{t_d,x}$ has to be best responding to the equilibrium,

$$\sigma_d^{t_d,x} > 0 \Rightarrow U_d^{t_d}(\hat{f}_d^x, \hat{f}_p^x) \geq x > 1 - w/i > 0.$$

Thus,

$$\left[\exists t_p \in \{s, w\} : \hat{f}_d^x(t_d|t_p) > 0\right] \Rightarrow U_d^{t_d}(\hat{f}_d^x, \hat{f}_p^x) > 1 - w/i > 0. \tag{20}$$

It follows that, for any t_d whose bid has positive mass in the criminal trial (i.e., $\sigma_d^{t_d,x} > 0$), $\phi_d^{t_d,x}$ has no atom at the zero bid. Otherwise, for any $t_p \in \{s, w\}$, the zero bid is an atom of the cdf $\Phi_d(\cdot|t_p)$ of the bids submitted by player d from the viewpoint of the opponent type (p, t_p) ($\Phi_d(\cdot|t_p)$ defined in (9)). Then by equilibrium conditions, the zero bid is not an atom of ϕ_p^s or ϕ_p^w, and hence by the definition of $\Phi_d(\cdot|t_d)$, it is not an atom of $\Phi_d(\cdot|t_d)$ $(\forall t_d)$ either. Consequently, for any t_d such that $\sigma_d^{t_d,x} > 0$ and the support of his bid distribution $\phi_d^{t_d,x}$ contains zero (such zero-bidding t_d exists due to Lemma 6), t_d gets only zero surplus in the trial, contradicting (20).

Consequently, for any $t_p \in \{s, w\}$ that contribute a positive mass of bids in the criminal trial and whose bid support contains zero (again such zero-bidding t_p exists due to Lemma 6), the surplus from the criminal trial is zero. In other words, $U_p^{t_p}(\hat{f}_d^x, \hat{f}_p^x) = 0$ for at least one $t_p \in \{s, w\}$. Then the maximum bid $\overline{b}$ in $(\phi_p^{s,x}, \phi_p^{w,x}, \phi_d^{i,x}, \phi_d^{g,x})$ cannot be less than w. Otherwise, (p, t_p) can bid $\overline{b}$ to have a positive expected payoff $1 - \overline{b}/t_p > 1 - w/t_p \geq 0$, as $t_p \geq w$. Now that $\overline{b} \geq w$, the type t_d of the defendant who has $\overline{b}$ in the support of his bidding strategy $\phi_d^{t_d,x}$ has the equilibrium surplus $1 - \overline{b}/t_d \leq 1 - w/t_d \leq 1 - w/i$, with the last inequality due to $i \geq t_d \geq g$. But this contradicts (20). The claim is thus proved.

By the claim just proved, $\tilde{U}_p^{t_p}(x) = 1 - x$ for each $t_p \in \{s, w\}$. Then (19) implies that

$$f(g|t_p)s/g \geq 1 - x.$$

for each $t_p \in \{s, w\}$. This being true for all $x > 1 - w/i$, we have

$$f(g|t_p)s/g \geq 1 - (1 - w/i) = w/i$$

for each $t_p \in \{s, w\}$. Thus follows the conclusion of the lemma. ■

A.5 A Punishing Continuation Equilibrium to the Prosecutor

Lemma 9 *For any posterior belief $\hat{f}_d := (\hat{f}(\cdot|s), \hat{f}(\cdot|w))$ about the defendant's type, the all-pay auction given belief system $(\hat{f}_d, \hat{f}_{p=s})$ ($\hat{f}_{p=s}(s|t_d) = 1\ \forall t_d$) has a unique BNE and, if $\hat{f}(i|s) \geq \frac{1/g-1/s}{1/g-1/i}$, then*

$$\begin{aligned} U_p^s(\hat{f}_d, \hat{f}_{p=s}) &= 0 \\ U_p^w(\hat{f}_d, \hat{f}_{p=s}) &= 0 \quad \text{if } \hat{f}_d = f_d \\ U_d^i(\hat{f}_d, \hat{f}_{p=s}) &= 1 - s/i \\ U_d^g(\hat{f}_d, \hat{f}_{p=s}) &= 1 - \frac{s\hat{f}(i|s)}{i} - \frac{s\hat{f}(g|s)}{g}. \end{aligned}$$

Proof Within the statement of Lemma 7, conduct the following swapping of the roles between the two players and their corresponding types:

$$[d \leftrightarrow p, s \leftrightarrow i, w \leftrightarrow g]\,.$$

Part (ii) of the lemma after the role swap becomes the current lemma.[24] ■

A.6 Proposition 3

The proof has been outlined in the main text. The details are provided by the following claims, each under the conditions assumed in the proposition.

Claim A $U_p^s(f_d, \hat{f}_{p=s}) = U_p^w(f_d, \hat{f}_{p=s}) = 0$.

Proof Condition (ii) in the proposition, coupled with the $s < i$ in Condition (i), implies $s < g$. Then the assumption $f(i|s) \geq (1/g - 1/s)/(1/g - 1/i)$ in Lemma 9 holds trivially, which implies this claim. ■

Claim B $U_d^{t_d}(f_d, \hat{f}_{p=s}) \geq x_*$ *for each* $t_d \in \{i, g\}$.

Proof Condition (i) of the proposition says $s/i < 1$, which implies by (5) that $1 < (1/w - 1/i)/(1/w - 1/s)$. This, coupled with the other part $f(s|i) < s/i$ of (i), implies that Part (iii)

[24] The counterparts to the other parts of Lemma 7 are not needed in my proofs and hence omitted.

of Lemma 7 applies. Thus,

$$U_d^i(\hat{f}_{d=i}, f_p) = f(w|i) - \frac{w}{i}\left(1 - \frac{if(s|i)}{s}\right) \tag{21}$$

$$U_d^g(\hat{f}_{d=i}, f_p) = \max\left\{ \begin{array}{l} f(w|g)\left(1 - \frac{1}{if(w|i)} w\left(1 - \frac{if(s|i)}{s}\right)\right), \\ 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) \end{array} \right\}. \tag{22}$$

Plug (22) into the construction $x_* = U_d^g(\hat{f}_{d=i}, f_p)$ for the proposed PBE to obtain

$$x_* = \max\left\{ \begin{array}{l} f(w|g)\left(1 - \frac{1}{if(w|i)} w\left(1 - \frac{if(s|i)}{s}\right)\right), \\ 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) \end{array} \right\}.$$

To calculate $U_d^{t_d}(f_d, \hat{f}_{p=s})$, use Lemma 9 (which applies as explained in the proof of Claim A):

$$U_d^i(f_d, \hat{f}_{p=s}) = 1 - \tfrac{s}{i}$$
$$U_d^g(f_d, \hat{f}_{p=s}) = 1 - \tfrac{sf(i|s)}{i} - \tfrac{sf(g|s)}{g}.$$

Note

$$U_d^g(f_d, \hat{f}_{p=s}) = 1 - \frac{s}{i}(1 - f(g|s)) - \frac{s}{g}f(g|s) = 1 - \frac{s}{i} - sf(g|s)\underbrace{\left(\frac{1}{g} - \frac{1}{i}\right)}_{>0} < 1 - \frac{s}{i} = U_d^i(f_d, \hat{f}_{p=s}).$$

Thus, the claim $U_d^{t_d}(f_d, \hat{f}_{p=s}) \geq x_*$ for each $t_d \in \{i, g\}$ is the same as

$$1 - \frac{s}{i} - sf(g|s)\left(\frac{1}{g} - \frac{1}{i}\right) \geq \max\left\{ \begin{array}{l} f(w|g)\left(1 - \frac{1}{if(w|i)} w\left(1 - \frac{if(s|i)}{s}\right)\right), \\ 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) \end{array} \right\}. \tag{23}$$

To simplify the right-hand side, note

$$\begin{aligned} 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) &= 1 - \frac{w}{g} - \frac{if(s|i)}{g}\left(1 - \frac{w}{s}\right) \\ &= 1 - \frac{w}{g} - \left(\frac{i}{g} - \frac{i}{g}f(w|i)\right)\left(1 - \frac{w}{s}\right) \\ &= \left(1 - \frac{w}{s}\right)\frac{i}{g}f(w|i) + 1 - \frac{w}{g} - \left(1 - \frac{w}{s}\right)\frac{i}{g}. \end{aligned}$$

And

$$\begin{aligned} f(w|g)\left(1 - \frac{1}{if(w|i)} w\left(1 - \frac{if(s|i)}{s}\right)\right) &= f(w|g)\left(1 - \frac{1}{if(w|i)} w\left(1 - \frac{i}{s} + \frac{i}{s}f(w|i)\right)\right) \\ &= f(w|g)\left(1 - \frac{w}{s} + \frac{w}{f(w|i)}\left(\frac{1}{s} - \frac{1}{i}\right)\right) \\ &= f(w|g)\left(1 - \frac{w}{s}\right) + \frac{wf(w|g)}{f(w|i)}\left(\frac{1}{s} - \frac{1}{i}\right). \end{aligned}$$

Let

$$\Delta := 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) - \left(f(w|g)\left(1 - \frac{1}{if(w|i)}w\left(1 - \frac{if(s|i)}{s}\right)\right)\right).$$

By the previously displayed calculations,

$$\begin{aligned}
\Delta &> \left(1-\frac{w}{s}\right)\frac{i}{g}f(w|i) + 1 - \frac{w}{g} - \left(1-\frac{w}{s}\right)\frac{i}{g} - f(w|g)\left(1-\frac{w}{s}\right) - \frac{wf(w|g)}{f(w|i)}\left(\frac{1}{s}-\frac{1}{i}\right) \\
&= \left(1-\frac{w}{s}\right)\left(\frac{i}{g}f(w|i) - f(w|g) - \frac{i}{g}\right) + 1 - \frac{w}{g} - \frac{wf(w|g)}{f(w|i)}\left(\frac{1}{s}-\frac{1}{i}\right) \\
&\overset{(4)}{>} \left(1-\frac{w}{s}\right)\left(\frac{1}{g}\left(gf(w|g) + i - g\right) - f(w|g) - \frac{i}{g}\right) + 1 - \frac{w}{g} - \frac{wf(w|g)}{f(w|i)}\left(\frac{1}{s}-\frac{1}{i}\right) \\
&= \left(1-\frac{w}{s}\right)(-1) + 1 - \frac{w}{g} - \frac{wf(w|g)}{f(w|i)}\left(\frac{1}{s}-\frac{1}{i}\right) \\
&= w\left(\frac{1}{s}-\frac{1}{g}\right) - \frac{wf(w|g)}{f(w|i)}\left(\frac{1}{s}-\frac{1}{i}\right) \\
&\geq 0,
\end{aligned}$$

with the last line due to Condition (ii) of the proposition. Thus, the right-hand side of (23) is equal to the second item inside $\{\cdots\}$. In other words,

$$x_* = 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right), \tag{24}$$

and the claim follows from

$$\begin{aligned}
& 1 - \frac{s}{i} - sf(g|s)\left(\frac{1}{g}-\frac{1}{i}\right) \geq 1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right) \\
\iff\ & \frac{s}{i} + sf(g|s)\left(\frac{1}{g}-\frac{1}{i}\right) \leq \frac{w}{g} + \frac{i}{g}\left(1-\frac{w}{s}\right)f(s|i) \\
\impliedby\ & \frac{s}{i} + sf(g|s)\left(\frac{1}{g}-\frac{1}{i}\right) \leq \frac{w}{g} \\
\iff\ & f(g|s) \leq \frac{w/g - s/i}{s(1/g - 1/i)} = \frac{(1/g)(w/s) - 1/i}{1/g - 1/i},
\end{aligned}$$

with the inequality on the last line being Condition (iii) in the proposition. ■

Claim C $f(g|w)(1 - x_*) \geq 1 - U_d^i(\hat{f}_{d=i}, f_p)$.

Proof By (21) and (24), this claim is the same as

$$f(g|w)\left(1 - \left(1 - \frac{1}{g}\left(w\left(1 - \frac{if(s|i)}{s}\right) + if(s|i)\right)\right)\right) \geq 1 - \left(f(w|i) - \frac{w}{i}\left(1 - \frac{if(s|i)}{s}\right)\right).$$

Rewrite both sides of this inequality to see that it is the same as

$$f(g|w)\left(1-\left(1-\frac{w}{g}-\frac{if(s|i)}{g}\left(1-\frac{w}{s}\right)\right)\right)\geq 1-f(w|i)+\frac{w}{i}-\frac{w}{s}f(s|i),$$

namely,

$$f(g|w)\left(\frac{w}{g}+\frac{if(s|i)}{g}\left(1-\frac{w}{s}\right)\right)\geq \frac{w}{i}+f(s|i)\left(1-\frac{w}{s}\right).$$

Rewrite the inequality in its equivalent forms:

$$\begin{aligned}
& f(g|w)\frac{w}{g}+f(g|w)\frac{i}{g}f(s|i)\left(1-\frac{w}{s}\right)-f(s|i)\left(1-\frac{w}{s}\right)-\frac{w}{i}\geq 0\\
\iff\quad & w\left(\frac{f(g|w)}{g}-\frac{1}{i}\right)+f(s|i)\left(1-\frac{w}{s}\right)\left(f(g|w)\frac{i}{g}-1\right)\geq 0\\
\iff\quad & \left(\frac{f(g|w)}{g}-\frac{1}{i}\right)\left(w+f(s|i)\left(1-\frac{w}{s}\right)i\right)\geq 0.
\end{aligned}$$

Since $w+f(s|i)\left(1-\frac{w}{s}\right)i>0$, the desired inequality holds if $f(g|w)/g-1/i\geq 0$, which is Condition (iv) in the proposition. ■

The next claim is referred to in the main text, though not needed for the proposition.

Claim D *If* $f(g|s)=((1/g)(w/s)-1/i)/(1/g-1/i)$ *and* $f(g|w)=g/i$, *then*

$$sf(i|s)<wf(i|w)\iff 1-\frac{w}{s}<\frac{w}{s}g^2\left(\frac{1}{g}-\frac{1}{i}\right)^2,$$

and $sf(i|s)<wf(i|w)$ *holds if* w/s *is sufficiently close to one, or* $i-g$ *sufficiently large.*

Proof Note $sf(i|s)<wf(i|w)\iff sf(g|s)>wf(g|w)+s-w$. Plug in the values of $f(g|s)$ and $f(g|w)$ in the premise of the claim to obtain

$$\begin{aligned}
sf(i|s)<wf(i|w)\quad & \iff\quad s\cdot\frac{(1/g)(w/s)-1/i}{1/g-1/i}>w\cdot\frac{g}{i}+s-w\\
& \iff\quad s\left(\frac{(1/g)(w/s)-1/i}{1/g-1/i}-1\right)>-w\left(1-\frac{g}{i}\right)\\
& \iff\quad s\cdot\frac{(1/g)(w/s-1)}{1/g-1/i}>-w\left(1-\frac{g}{i}\right)\\
& \iff\quad \frac{s}{g}\cdot\frac{1-w/s}{1/g-1/i}<wg\left(\frac{1}{g}-\frac{1}{i}\right)\\
& \iff\quad 1-\frac{w}{s}<\frac{w}{s}g^2\left(\frac{1}{g}-\frac{1}{i}\right)^2.
\end{aligned}$$

Note that the inequality on the last line holds if w/s is sufficiently close to one, or $1/g-1/i$ sufficiently large. ■

A.7 Proposition 4

Condition (i) in this proposition is exactly the condition for Part (iii) of Lemma 7. The lemma therefore implies $U_d^i(\hat{f}_{d=i}, f_p) > U_d^g(\hat{f}_{d=i}, f_p) > 0$. Thus, $S := \left[U_d^g(\hat{f}_{d=i}, f_p), U_d^i(\hat{f}_{d=i}, f_p)\right]$ is nonempty. Let us verify that in the modified plea bargain game given this S there exists a separating PBE in which both types of the prosecutor offers $x_* := U_d^g(\hat{f}_{d=i}, f_p)$. Clearly this x_* satisfies the defendant's incentive condition, Ineq. (6). Given the restriction to S, the prosecutor's incentive condition, Ineq. (7), is reduced to

$$f(g|t_p)(1 - x_*) + f(i|t_p)U_p^{t_p}(\hat{f}_{d=i}, f_p) \geq \tilde{U}_p^{t_p}(0)$$

for each $t_p \in \{s, w\}$, with the right-hand side denoting the prosecutor's surplus from the criminal trial that ensues immediately after she deviates to $x = 0$ instead of offering any element in S. To satisfy this inequality, let the continuation equilibrium conditional on this deviation be the BNE given posterior system $(f_d, \hat{f}_{d=s})$, where $\hat{f}_{d=s}$ denotes the off-path posterior that the prosecutor's type is s for sure. By Condition (ii) in this proposition, Lemma 9 implies that $U_p^s(\hat{f}_d, \hat{f}_{p=s}) = U_p^w(\hat{f}_d, \hat{f}_{p=s}) = 0$ and hence $\tilde{U}_p^s(0) = \tilde{U}_p^w(0) = 0$. Hence the above inequality holds, and hence follows the prosecutor's incentive condition. ■

References

[1] Richard P. Adelstein. The plea bargain in theory: A behavioral model of the negotiated guilty plea. *Southern Economic Journal*, 44(3):488–503, 1978.

[2] Scott Baker and Claudio Mezzetti. Prosecutorial resources, plea bargaining, and the decision to go to trial. *Journal of Law, Economics, & Organization*, 17(1):149–167, 2001.

[3] Benjamin Balzer and Johannes Schneider. Managing a conflict: Optimal alternative dispute resolution. *RAND Journal of Economics*, 52(2):415–445, 2021.

[4] S. Banoyopadhyay and Bryan C. McCannon. The effect of the election of prosecutors on criminal trials. *Public Choice*, 161:141–156, 2014.

[5] Nicholas C. Bedard and Charles Z. Zheng. A contest without monotone equilibria. Mimeo, Department of Economics, The University of Western Ontario, July 16, 2019.

[6] Helmut Bester and Karl Wärneryd. Conflict and the social contract. *Scandinavian Journal of Economics*, 108:231–249, 2006.

[7] David Bjerk. 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, 2007.

[8] Richard T. Boylan. What do prosecutors maximize? evidence from the careers of U.S. attorneys. *American Law and Economics Review*, 7:379–402, 2005.

[9] Richard T. Boylan and Cheryl X. Long. Salaries, plea rates, and the career objectives of federal prosecutors. *Journal of Law and Economics*, 48:627–652, 2005.

[10] Kong-Pin Chen and Jue-Shyan Wang. Fee-shifting rules in litigation with contingency fees. *Journal of Law, Economics, & Organization*, 23(3):519–546, 2007.

[11] Olivier Compte and Philippe Jehiel. Veto constraint in mechanism design: Inefficiency with correlated types. *American Economic Journal: Microeconomics*, 1:182–206, 2009.

[12] Andrew F. Daughety and Jennifer F. Reinganum. Informal sactions on prosecutors and defendants and the disposition of criminal cases. *Journal of Law, Economics, & Organization*, 32:359–394, 2016.

[13] Andrew F. Daughety and Jennifer F. Reinganum. Settlement and trial: Selected analyses of the bargaining environment. In Francesco Parisi, editor, *The Oxford Handbook of Law and Economics*, volume 3, pages 229–246. Oxford University Press, 2017.

[14] Andrew F. Daughety and Jennifer F. Reinganum. Evidence suppression by prosecutor: Violations of the Brady rule. *Journal of Law, Economics, & Organization*, 34(3):475–510, 2018.

[15] Mark Fey and Kristopher W. Ramsay. Uncertainty and incentives in crisis bargaining: Game-free analysis of international conflict. *American Journal of Political Science*, 55:149–169, 2011.

[16] Ezra Friedman and Abraham L. Wickelgren. Chilling, settlement, and the accuracy of the legal process. *Journal of Law, Economics, & Organization*, 26(1):144–157, 2008.

[17] Samuel R Gross, Barbara O'Brien, Chen Hu, and Edward H. Kennedy. Rate of false conviction of criminal defendants who are sentenced to death. *Proceedings of the National Academy of Sciences - PNAS*, 111(20):7230–7235, 2014.

[18] Gene M. Grossman and Michael L. Katz. Plea bargaining and social welfare. *American Economic Review*, 73(4):740–757, 1983.

[19] Johannes Hörner, Massimo Morelli, and Francesco Squintani. Mediation and peace. *Review of Economic Studies*, 82:1483–1501, 2015.

[20] Ali Kamranzadeh and Charles Z. Zheng. Unequal peace. *International Economic Review*, 66(1):223–258, February 2025.

[21] William M. Landes. An economic analysis of the courts. *Journal of Law and Economics*, 14:61–108, April 1971.

[22] Jingfeng Lu, Zongwei Lu, and Christian Riis. Peace through bribing. Working Paper, arXiv:2107.11575 [econ.TH], April 2023.

[23] Bryan C. McCannon. Prosecutor elections, mistakes, and appeals. *Journal of Empirical Legal Studies*, 10:696–714, 2013.

[24] Roger B. Myerson. Mechanism design by an informed principal. *Econometrica*, 51(6):1767–1797, November 1983.

[25] Jennifer F. Reinganum. Plea bargaining and prosecutorial discretion. *American Economic Review*, 78(4):713–728, 1988.

[26] Lucas Rentschler and Theodore L. Turocy. Two-bidder all-pay auctions with interdependent valuations, including the highly competitive case. *Journal of Economic Theory*, 163:435–466, 2016.

[27] David Rosenberg and Kathryn E. Spier. Incentives to invest in litigation and the superiority of the class action. *Journal of Legal Analysis*, 6(2):305–365, 2014.

[28] Henk Schouten. *Essays on the All-Pay Auction*. PhD thesis, The University of Western Ontario, London, Ontario, 2022.

[29] Ron Siegel. Asymmetric all-pay auctions with interdependent valuations. *Journal of Economic Theory*, 153:684–702, 2014.

[30] Ron Siegel and Bruno Strulovici. Judicial mechanism design. *American Economic Journal: Microeconomics*, 15(3):243–270, 2023.

[31] Bernardo S. Silveira. Bargaining with asymmetric information: An empirical study of plea negotiations. *Econometrica*, 85(2):419–452, March, 2017.

[32] Kathryn E. Spier. Pretrial bargaining and the design of fee-shifting rules. *The RAND Journal of Economics*, 25(2):197–214, 1994.

[33] Charles Z. Zheng. Bidding collusion without passive updating. *Journal of Mathematical Economics*, 85:70–77, 2019.

[34] Charles Z. Zheng. Necessary and sufficient conditions for peace: Implementability versus security. *Journal of Economic Theory*, 180:135–166, March 2019.