Reducing Unjust Convictions:
Plea Bargaining, Trial, and Evidence Suppression/Disclosure *

by

Andrew F. Daughety and Jennifer F. Reinganum**
Department of Economics and Law School
Vanderbilt University

July 2018
revised April 2019

Abstract

We develop a dynamic model of a criminal case, from arrest through plea bargaining and (possibly) trial, allowing for the potential discovery of exculpatory evidence by prosecutors (who choose whether to disclose it) and defendants. We consider three regimes: (1) no disclosure required; (2) disclosure required before trial; and (3) disclosure required from the point of arrest onward. Both disclosure regimes are more (less) likely to convict the guilty (innocent) than no required disclosure, but more extensive disclosure comes at a cost of a higher (lower) likelihood of conviction of innocent (guilty) defendants when no exculpatory evidence has been discovered.

Keywords: Evidence suppression, disclosure, prosecutorial misconduct, plea bargaining

JEL CODES: K4, D82, D73

* We thank Eric Talley, Jeffrey Gordon, and Robert Scott for providing a stimulating research environment at Columbia Law School during the initial work on this paper. We also thank Justin Fox, Xinyu Hua, Nancy King, Joel Sobel, Kathy Spier, and participants in seminars at Vanderbilt University, the Law and Economics Theory 8 Conference, and the 2019 Winter Meetings of the NBER Law and Economics Program for comments on a previous version. We further thank four anonymous referees and the Editor for extensive and helpful comments and suggestions.

** andrew.f.daughety@vanderbilt.edu; jennifer.f.reinganum@vanderbilt.edu

1. Introduction

As Robert H. Jackson (1940) once famously observed, “The prosecutor has more control over life, liberty, and reputation, than any other person in America.”[1] Prosecutors often supervise investigative officials (including police) and may actively engage in criminal investigation. They choose which cases to pursue, which charges to bring against a defendant, and what sentences to offer in lieu of trial (plea offers). A significant degree of independence, and the availability of state resources (police, other investigative services including crime labs, and legal authority), encourage prosecutorial zeal to seek justice for victims. On the other hand, abuse of power by prosecutors (possibly reflecting pursuit of enhanced career opportunities) can have stunning consequences for defendants, particularly for those who are actually innocent.[2] For example, it is the prosecutor who brings capital charges if allowed, or who proposes (say) lifetime sentences in other jurisdictions. Problems of abuse of power, and how to incentivize prosecutors to act in society’s interests, as well as finding ways to restrain some of their power, have been a continuing concern for society. Standard decentralized tools to influence individual agent behavior, such as civil liability, are generally unavailable in this context,[3] so regulatory approaches have been relied upon. Over the last sixty years, courts have attempted to address the problem of limiting prosecutorial abuse of power (without overly diminishing desirable prosecutorial zeal), particularly via rules requiring disclosure

[1] Jackson should know: he was U.S. Solicitor General, U.S. Attorney General, and an Associate Justice of the U.S. Supreme Court. In 1945 he took a leave from the Court to assume the role of Chief U.S. Prosecutor at the Nuremberg Trials.

[2] Official misconduct is not the only reason for unjust convictions, but according to the National Registry of Exonerations, it is a contributing factor in 52% of the cases in their database of 2238 exonerations since 1989. Other common reasons include witness misidentification (29%); perjury or false accusation (57%); false or misleading forensic evidence (24%); and false confession (13%). Note that more than one factor can be a contributing basis in the conviction of an innocent defendant. http://www.law.umich.edu/special/exoneration/Pages/ExonerationsContribFactorsByCrime.aspx last accessed 6/19/18.

[3] Prosecutors are absolutely immune from civil liability for misconduct in advocative roles. See Gershman (2015, Section 14:14) for an extensive elucidation of this. The U.S. Supreme Court has required suits for damages against municipalities to demonstrate what amounts to a pattern of deliberate indifference to constitutional rights; for example, see the majority opinion in *Connick v. Thompson* (2011).

by the prosecutor of exculpatory and impeachment evidence.[4] For instance, the *Brady* rule[5] requires the prosecution to disclose exculpatory evidence to the defense before trial. Dissatisfaction with the effectiveness of court-supplied rules has caused some state legislatures to require disclosure of all evidence upon receipt ("open-files"). Turner and Redlich (2016, pp. 302-305) classify the federal system and 10 states as following a "closed files" policy; 17 states follow an "open files" policy; and the remaining states are somewhere in between.[6] The degree of reliance on *Brady* or the use of other disclosure requirements varies from state-to-state, including determination of when disclosure must occur (and what material must be disclosed; see Hooper et. al., 2004).

In this paper we develop a dynamic model of the criminal justice process so as to understand the tradeoffs and implications of various evidence-disclosure requirements. We consider three stylized disclosure regimes: (1) no required disclosure (denoted as N); (2) disclosure of exculpatory evidence before a trial (denoted as B, for the *Brady* rule); and (3) disclosure of exculpatory evidence from the point of arrest until disposition of a case (denoted as X, for extensive disclosure).

Data from the U. S. Department of Justice reveal that 88% of all federal defendants choose to conclude a plea offer rather than go to trial and approximately 97% of all federal convictions are

[4] Exculpatory evidence is evidence that supports a defendant's assertion of innocence of a crime. Inculpatory evidence is evidence that supports a prosecutor's assertion that a defendant is guilty of a crime. Impeachment evidence is evidence that undermines, for example, the testimony of a witness for the prosecution. Thus, exculpatory evidence concerns the strength of the defendant's case while impeachment evidence concerns the weakness of the prosecution's case.

[5] *Brady v. Maryland*, 1963, which will be discussed in more detail in Section 2.

[6] See Grunwald (2017) for a detailed empirical examination of open-file discovery in North Carolina and Texas and see Turner and Redlich (2016) for a detailed comparison of North Carolina with Virginia (which is a "closed files" state). Grunwald finds that there are more motions to suppress under open files, from which he infers that more material is being disclosed. He finds no other significant effects, but his data is limited to yearly observations (aggregated to the state level), which may account for the weak statistical results. Turner and Redlich surveyed chief prosecutors and defense attorneys (public defenders and private attorneys) in the two states, and found that "North Carolina defenders stated that they received ... exculpatory and impeachment evidence more frequently than did their Virginia counterparts." and that "... responses may be read to support the hypothesis that open-file discovery produces more consistent disclosure of *Brady* material." (p. 331).

due to plea bargains.[7] To date, no Supreme Court decision has determined that *Brady* applies to plea bargaining (and not just trial),[8] which means that the vast share of defendants (since 88% of the total number plead guilty) are generally not protected by the primary existing disclosure rule. This is why we consider the third disclosure rule, denoted above as X, as that would provide for disclosure of any exculpatory evidence before plea bargaining. We show that changes in the timing of disclosure (that is, shifts among regimes N, B, and X) have important distributional effects not only between innocent and guilty defendants, but also between subgroups of innocent defendants. These effects arise due to bargaining between prosecutors and defendants in the shadow of the disclosure requirement for prosecutors. In particular, we find that requiring earlier evidence disclosure (as compared to disclosure just before trial) may be <u>disadvantageous</u> for innocent defendants that face a prosecutor who does not possess exculpatory evidence.

From the perspective of economic models of disclosure (see Section 2), our model reflects a setting wherein the prosecutor may, or may not, possess the evidence of interest, so that the possession of evidence is, itself, private information for the prosecutor. We will refer to a prosecutor who possesses exculpatory evidence as "informed" and to one who doesn't as "uninformed." This means that the classic "unraveling result" (wherein the possessor of private information always discloses it in equilibrium) need not occur because it is not common knowledge that the prosecutor possesses such information. Furthermore, plea offers made by a prosecutor need not reveal whether

[7] Estimates at the federal level are for the last decade and are based on caseload statistics tables (Table D-4, U.S. District Courts - Criminal Defendants Terminated, by Type of Disposition and Offense; last accessed 07/09/18) at http://www.uscourts.gov/report-names/federal-judicial-caseload-statistics. There appears to be no readily-available, systematic data for each of the fifty states.

[8] In *U.S. v. Ruiz* (2002), a unanimous Court ruled that a right to exculpatory <u>impeachment</u> evidence prior to entering a plea agreement is not guaranteed by the Constitution (and, hence, by *Brady*). Writing for the Court, Justice Stephen Breyer stated that: " ... a defendant who pleads guilty forgoes a fair trial as well as various other accompanying constitutional guarantees." Note that this pertains to the disclosure of a potential weakness of the prosecution's case (e.g., impeachment of a witness for the prosecution), not the strength of the defendant's case (i.e., evidence of innocence).

the prosecutor possesses exculpatory evidence. We find that in regimes N and X there is a unique equilibrium plea offer (which is revealing) while regime B only has pooling equilibria (a refinement selects one of these). We characterize the differences between the regime-specific offers and we find the thresholds that characterize accept/reject behavior by defendants. Furthermore, our model (unlike the vast majority of models in the plea-bargaining literature) shows why some innocent defendants will (in equilibrium) accept the equilibrium plea offer, which is consistent with real-world exoneration data (see footnote 2). We use our model to derive *ex ante* preferences over the regimes for prosecutors, innocent defendants, and guilty defendants. Finally, we use the foregoing results to derive the regime-specific probabilities of conviction for: (1) innocent defendants facing prosecutors who are suppressing exculpatory evidence; (2) innocent defendants facing prosecutors who have <u>not</u> obtained exculpatory evidence; and (3) guilty defendants. These probabilities are then used to compare the three disclosure regimes for these three groups of defendants.

1.1. Plan of the Paper and Overview of Results

In Section 2 we provide a discussion of the relevant institutional and legal background, and then turn to a discussion of the related literature on disclosure and on plea bargaining. Section 3 provides the basic setup of our three-period dynamic model, which starts with the arrest of a defendant and proceeds through plea bargaining and possible trial, allowing for the prosecutor to drop the case at various points along the way (including before or after plea bargaining). The defendant's acceptance of a plea offer, or conviction at trial, results in incarceration, but for innocent defendants the possibility of exoneration occurs with positive probability, both before and after conviction (which has been important in some real-world cases). Section 4 provides the analysis of the equilibrium actions of the prosecutor and the defendant under each of the three disclosure regimes. A key result is that in regimes N and X, a prosecutor has a credible threat to go to trial following rejection of the plea offer, whereas in regime B an informed prosecutor makes the same plea offer as an uninformed prosecutor, but drops the case if the offer is rejected. This, in turn, makes the defendant less willing to accept any given offer in regime B. At the end of Section 4 we compare the behavior by the prosecutor and the defendant in each disclosure regime and discuss agents' preferences among the regimes.

Section 5 discusses our notion of unjust and just convictions. An innocent defendant who is convicted (either via accepting a plea offer or via trial) when the prosecutor actually possesses exculpatory evidence suffers an "informed unjust conviction." An innocent defendant who is convicted (either via accepting a plea offer or via trial) when the prosecutor actually does <u>not</u> possess exculpatory evidence suffers an "uninformed unjust conviction." Finally, a guilty defendant who is convicted (either via accepting a plea offer or via trial) is viewed as a "just conviction." Section 5 derives the probabilities of these three types of conviction for each of the three regimes, which allows us to compare the impact of the different disclosure policies on unjust and just convictions. We show that both regime B and regime X entail (sometimes weakly) lower probabilities of an unjust conviction (both informed and uninformed) and a (sometimes weakly) higher probability of a just conviction than would be true under the no-disclosure requirement (i.e., regime N). However, the yet earlier disclosure associated with regime X need not be "better" than that associated with regime B from society's perspective. While regime X reduces the likelihood of an informed unjust conviction (in comparison with regime B), the reverse holds for uninformed unjust convictions: the probability of uninformed unjust convictions under regime X is higher than under regime B because D is more willing to accept under regime X than under regime B.

Moreover, regime X leads to a decrease in the probability of just convictions in comparison with regime B. Thus, the social choice between regimes B and X involves both a classical tradeoff between convicting innocents versus acquitting/releasing the guilty, as well as a new tradeoff between those innocents whose conviction reflects prosecutor abuse of power and those whose conviction does not. Section 6 provides a summary and a discussion of our analysis.

2. Institutional Background and Review of Relevant Theoretical and Empirical Literature

2.1. Institutional Background

In 1963 the U.S. Supreme Court ruled in *Brady v. Maryland* that a prosecutor must disclose exculpatory evidence favorable to a defendant that is "material" to guilt or punishment, where evidence is material if its disclosure could change the outcome of a trial.[9] Failure to disclose material exculpatory evidence is a violation of the defendant's constitutional right to due process. Over time, however, implementation of this right has been imperfect: an authority on prosecutorial misconduct has observed that "... violations of *Brady* are the most recurring and pervasive of all constitutional procedural violations, with disastrous consequences ..." (Gershman, 2007, p. 533).

Two points are worth stressing. First, the decision regarding what evidence is *material* to the defense is made by the prosecutor and is thus potentially subject to both cognitive bias and strategic manipulation. In an attempt to alleviate concerns that prosecutors under-disclose due to self-serving decisions about what is "material," several individual states have adopted versions of "open-file discovery." The basic idea of an open file is that whatever evidence the prosecution has

[9] This was subsequently extended (through a series of further judicial decisions) to include: (1) evidence that can be used to impeach a witness; (2) evidence favorable to the defense that is in the possession of the police; and (3) evidence that the prosecution knew (or should have known) that their case included perjured testimony (see Kozinski, 2015, and Keenan, et. al., 2011).

in its possession, that evidence should be promptly disclosed to the defense as well.[10]

Second, particularly since the Court's unanimous decision in *Ruiz*, the *Brady* requirement is generally interpreted as only pertaining to evidence to be used at trial and specifically not to plea bargaining (this despite the fact that *Ruiz* is focused on impeachment evidence). Using the estimates from Section 1, this means that up to 88% of all federal defendants chose to accept a plea offer in lieu of trial without the assurance of access to all relevant evidence in the case.

2.2. Related Literature

This paper is related to two strands of literature in economics and in law and economics. First, there is a substantial literature on incentives for agents (usually sellers in a market) to disclose their private information (usually about product quality; see Dranove and Jin, 2010, for a recent survey on the disclosure of product quality). When disclosure is costless and it is common knowledge that the seller is informed, then "unraveling" occurs: all private information is disclosed (see Grossman, 1981, and Milgrom, 1981). When disclosure is costly (see Jovanovic, 1982, and Daughety and Reinganum, 2008) or there is a chance that the agent has no private information (see Dye, 2017, and Daughety and Reinganum, 2018), then complete unraveling does not occur.[11]

In Daughety and Reinganum (2018) we develop a model of prosecutors taking cases to trial (that is, abstracting from plea bargaining) in order to address the incentives for the suppression of

[10] There are variations among what the states require to be in the file, and some types of evidence are more closely-held than others. Important practical considerations may limit the disclosure of some information. For example, disclosure of witness addresses can be delayed until the day of testimony (see *U.S. v Higgs*, 1983). Mosteller (2008) proposes an expansive policy ("full open-file discovery," see p. 263) that would include not only evidence in possession of the prosecution but also evidence in the possession of the police or other investigative agencies.

[11] Matthews and Postlewaite (1985) and Shavell (1994) show that, when an agent chooses whether to acquire and disclose information, then a rule mandating disclosure can discourage information acquisition (as compared to voluntary disclosure). Garoupa and Rizzolli (2011) interpret this finding in the context of the *Brady* rule. They argue that, if disclosure of exculpatory evidence is mandatory, then a prosecutor with a putative strong case may curtail her search for additional evidence, which could harm an innocent defendant.

exculpatory evidence in the shadow of the *Brady* rule. A prosecutor may (or may not) have observed exculpatory evidence. She has a utility for winning a case (career concerns), but also experiences a moral disutility for knowingly convicting an innocent defendant. Prosecutors are heterogeneous with respect to this disutility, which is their private information. A convicted (but innocent) defendant may later discover exculpatory evidence and a judge will then void the conviction and choose whether to order an investigation, which results in a penalty if suppression is verified. Judges are heterogeneous in their opportunity costs of pursuing suspected misconduct (which is each judge's private information). In equilibrium, some prosecutors suppress exculpatory evidence and some judges investigate and punish suspected misconduct. We also consider teams of prosecutors who can self-organize the flow of information within the team. We show that this results in the compartmentalization of authority regarding the receipt and control of exculpatory evidence.

The second strand of related literature is about plea bargaining; these models involve varying degrees of prosecutor concern for the innocent. For instance, Landes (1971) provides a complete information model wherein there are no innocent defendants; hence the assumption that the prosecutor maximizes expected sentences is used. Subsequent papers by Grossman and Katz (1983) and Reinganum (1988) assumed that prosecutors suffered some disutility from convicting an innocent defendant, but were committed to taking a case to trial if plea bargaining failed.

Nalebuff (1987) showed that, if a prosecutor with concerns about convicting innocent defendants were not committed to trial, then a credibility constraint must be included;[12] that is, the prosecutor's offer must induce sufficient rejection of the plea offer among guilty defendants so that

[12] Nalebuff was actually modeling a civil trial, so the "prosecutor" was a civil plaintiff, but the model translates fairly directly to a criminal trial.

trial remains a credible threat. Several articles have explored various aspects of plea bargaining while incorporating such a credibility constraint; see Franzoni (1999), Baker and Mezzetti (2001), Bjerk (2007; where the credibility constraint pertains to a jury rather than the prosecutor), and Daughety and Reinganum (2016). Finally, there is also a related literature on civil suits with negative expected value (NEV). We are aware of only one model with NEV suits wherein the informed plaintiff makes the settlement demand (Farmer and Pecorino, 2007).[13] If the plaintiff's suit has NEV, then she will drop it if the defendant rejects her settlement demand. Anticipating that some plaintiffs will drop their suits following rejection, the defendant rejects any given settlement demand more often. The behavior of our innocent defendant in regime B follows from the same logic.

Much of the previous work on plea bargaining wherein D has private information about his guilt or innocence focuses on using plea bargaining to screen innocent and guilty defendants. These models typically feature a plea offer that innocent defendants always reject, whereas guilty defendants accept this offer with positive probability.[14] This result arises because, except for the fact that an innocent defendant is less likely than a guilty defendant to be convicted at trial (e.g., the distribution of evidence against an innocent defendant is more favorable than that of the guilty defendant), the two types of defendant are assumed to be otherwise identical. In our model, the two

[13] Most of these contributions follow Bebchuk (1988) and Katz (1990) in employing a screening model wherein an uninformed defendant makes a settlement offer to an informed plaintiff, whose suit may have negative expected value. A more recent contribution involves multiple rounds of (screening) settlement offers, with an intervening discovery stage (Schwartz and Wickelgren, 2009).

[14] See Reinganum (1988) for a signaling model wherein both innocent and guilty defendants accept the plea offer with positive probability. Daughety and Reinganum (2016) also provide a model wherein some innocent defendants accept plea offers due to risk or ambiguity aversion. See Mungan and Klick (2016) who also employ heterogeneous attitudes towards risk to generate outcomes with both some innocent and some guilty defendants accepting a plea offer. Mungan and Klick argue that exoneree compensation can incentivize innocent defendants to reject an arbitrary plea offer more often while holding constant the acceptance rate of guilty defendants.

types of defendant are equally likely to be convicted at trial with the same evidence, but the innocent defendant may find exculpatory evidence before trial. This would lead to the same sorting effects as in previous models, except that we add another dimension of defendant heterogeneity which results in some innocent defendants also accepting the plea offer.

As noted above, models of prosecutors vary in the extent to which they include a disutility for convicting innocent defendants. Empirical work on this issue finds evidence of career concerns, but also justice-related concerns. For instance, Glaeser, Kessler, and Piehl (2000) find that some federal prosecutors appear to be motivated by reducing crime while others appear to be motivated primarily by career concerns. Boylan and Long (2005) find that higher private salaries are associated with assistant U.S. attorneys taking more cases to trial; this suggests that they might pursue more trial experience in anticipation of leaving for a well-paid private-sector job. Boylan (2005) finds that the length of prison sentences obtained by a U.S. attorney is positively related to positive outcomes in his or her career path. Bandyopadhyay and McCannon (2014) find that prosecutors subject to reelection pressure (i.e., chief prosecutors for an office) try to increase the number of convictions obtained through trial, and McCannon (2013) finds that this reelection pressure can lead to more wrongful convictions (based on more reversals on appeal).

Primarily for simplicity (as the model is otherwise rather complicated), in this paper we will assume that prosecutors do not suffer an internal moral disutility from knowingly convicting an innocent defendant.[15] However, since a prosecutor who suppresses exculpatory evidence and convicts an innocent defendant may be subject to a penalty (under regimes B or X), we will find that

[15] The model could accommodate a moral disutility that was, by itself, insufficient to generate disclosure, so that some version of the *Brady* rule would be required for disclosure before trial. As noted above, in Daughety and Reinganum (2018) we provided a model with heterogeneous disutility, and that feature was an important determinant of a prosecutor's compliance with the *Brady* rule. While we believe this could be a valuable extension of the current model, we think it is second-order relative to the issues explored in this paper, and therefore we do not incorporate it here.

there is still a credibility incentive that must be addressed, as a prosecutor who fails to obtain a conviction via plea bargain may want to drop the case before trial so as to avoid a penalty.

3. Model Setup and Notation

In this section, we will provide the notation and the primary model structure we employ in the rest of the paper, and describe the plea bargaining game between the prosecutor and the defendant. We envision this game being played over three periods of unequal length. At the beginning of period 1 (time zero), D is arrested and the police provide inculpatory evidence to P. On its own, the inculpatory evidence implies that P would win at trial with (a strictly fractional) probability π; this assessment is commonly known by P and D. Due to errors in the arrest process, D may be guilty (G) or innocent (I), which is D's private information.[16] Let $\lambda_0 \equiv \Pr\{D \text{ is } I\}$ and assume that this parameter is common knowledge to P and D.[17] Although D knows his type, it does not directly affect whether he is convicted at trial. Rather, the evidence that is presented at trial will determine that outcome; of course, D's type may affect the evidence that is presented at trial.

3.1. Utility functions for P and D

In what follows we use "T" as a subscript to denote "trial" and "O" (that is, "Oh") to denote "offer." For simplicity, we will assume that P's utility function is common knowledge, and that P's utility is S_T when D is convicted at trial (and receives the exogenously-determined sentence S_T), and P's utility is S_O when D accepts the plea offer S_O. However, we assume that D's disutility has a heterogeneous component associated with conviction. Thus, upon conviction at trial, D's disutility

[16] We will sometimes refer to the generic defendant as D. When we want to emphasize a specific type of D, we will use "D of type .." or simply refer to I or G. Note that the outcome of a trial will be acquittal or conviction, which can occur to either an innocent or a guilty D; similarly, a plea bargain (resulting in conviction) could be accepted by either type of D.

[17] Innocent Ds who possess evidence that clears them of the charges are assumed to have already provided that evidence and have had their cases dropped. Thus, λ_0 reflects innocent Ds who have not (yet) discovered such evidence.

is $S_T + \delta$, where δ is uniformly distributed on $[0, \delta^M]$ and is private information for D. Similarly, we assume that D's disutility upon accepting a plea offer of S_O is $S_O + \delta$, where δ is uniformly distributed on $[0, \delta^M]$ and is private information for D.[18] Thus, a complete description of D's type is (G, δ) or (I, δ).

This heterogeneous component of D's disutility can reflect follow-on losses from formal sanctions, such as the loss of the right to vote, as well as informal sanctions (from members of society) such as limitations on access to jobs, housing, and educational programs following a conviction.[19] Moreover, in this formulation a plea bargained sentence of $S_O = 0$ is not equivalent to a case being dropped (because D still bears the utility loss of δ). Again, this is plausible since a conviction is part of the record and may trigger the aforementioned additional losses.[20] Note that these utility functions are not type-dependent or event-dependent: (1) an informed P and an uninformed P obtain the same utility from a conviction at trial or an accepted plea (e.g., P only cares about whether D is G or I for the purposes of computing expected payoffs; P's utility function itself does not depend on whether D is innocent); (2) both the innocent and guilty types of D suffer equally when convicted, either via plea or trial; that is, for any given D, the same δ reflects the disutility of conviction (independent of whether it was due to accepting a plea bargain or being

[18] Note that S_T and S_O need not be interpreted literally as the "sentence" D receives upon conviction; the actual sentence is whatever generates the utility levels S_T and S_O for P. But it is convenient to refer to S_T and S_O as the sentences at trial and under a plea deal, respectively.

[19] In Daughety and Reinganum (2016) we consider how limited information about the plea bargaining process can influence the beliefs of outside observers (citizens), leading to the imposition of these sorts of informal sanctions by members of society on both P and D. In that article we also allow for defendants who may be risk averse or ambiguity averse. We abstract from those considerations in this paper, so as to obtain closed-form solutions.

[20] This is why we have employed an additive formulation of D's disutility for a formal sanction of S. With a multiplicative specification, a sanction of S = 0 will create a disutility of zero, even though D now has a criminal record and will suffer social losses.

convicted at trial).[21]

3.2. Timing and Information Structure

In this subsection we will discuss the timing and information structure of the model; the reader may wish to follow along with Figure 1, which summarizes the timing and information structure, and which appears at the end of Section 3. During period 1, P has the opportunity to (privately) observe a random variable, denoted θ_1, which represents the possible discovery by P of exculpatory evidence;[22] thus, in this analysis, P can be one of two types: informed or uninformed. As in Daughety and Reinganum (2018), we assume that exculpatory evidence E is "perfect" in the sense that: (1) if it is presented at trial, then D will be acquitted for sure; and (2) if D is G, then no such exculpatory evidence exists.[23] We assume that exculpatory evidence for an innocent D always exists: an innocent D knows that he did not commit the crime, and since a crime was committed, someone else committed the crime. Therefore some exculpatory evidence exists, though it may not be discovered by P or D. To keep the analysis manageable, we assume that discovery of exculpatory evidence by P can only occur in period 1 while discovery by D can only occur in periods 2 and 3.

If P observes exculpatory evidence, we denote that outcome as $\theta_1 = E_1$ (where the subscript 1 indicates that E was observed by P in period 1). If P does not observe exculpatory evidence in

[21] Some models assume that an innocent D suffers more from the same outcome than a guilty D, but we abstract from such arguments. One could allow such differentiation by drawing the relevant δ from a G- or I-type-dependent distribution; since D's guilt or innocence is his private information, all this would do is complicate the exposition. Here we assume the distribution of δ does not depend upon G or I.

[22] For convenience, we use the term "exculpatory evidence" to refer to both evidence which directly supports the defendant's assertion of innocence (it exculpates the defendant) as well as evidence which could be used to impeach a prosecution witness (impeachment evidence), but would not in and of itself prove D's innocence. For more detail on these two types of evidence, see Gershman (2015, Section 5.9).

[23] In reality, exculpatory evidence may not be perfect, so its presence may only reduce the chance of conviction. Furthermore, imperfect exculpatory evidence may be observed even though D is of type G. Consideration of imperfect exculpatory evidence would considerably complicate the model (injecting a variety of additional parameters and inference conditions), require a separate model of trial and evidence assessment, and distract from our focus on the dynamics of plea bargaining and trial under alternative disclosure regimes.

period 1, we denote that outcome as $\theta_1 = \varphi_1$. Let $\gamma \equiv \Pr\{\theta_1 = E_1 \mid D \text{ is } I\}$; that is, γ is the probability that P observes E in period 1, given that D is innocent. Thus, if P observes $\theta_1 = E_1$, then P knows that D is innocent, whereas if P observes $\theta_1 = \varphi_1$, then P knows that D could be either innocent or guilty. Thus, a P that observed $\theta_1 = E_1$ is an "informed P," and one that observed $\theta_1 = \varphi_1$ is an "uninformed P."[24] Let $\lambda_1 \equiv \Pr\{D \text{ is } I \mid \theta_1 = \varphi_1\}$ be P's posterior belief that D is innocent given that P observed no exculpatory evidence. By Bayes' Theorem, $\lambda_1 = \lambda_0(1 - \gamma)/[1 - \lambda_0 + \lambda_0(1 - \gamma)]$. Clearly this posterior assessment by P of D's innocence is lower than the prior λ_0. We think of P as forming this assessment during period 1. Also before the end of period 1, P makes a public report, denoted as r, about what she observed. If P observed $\theta_1 = E_1$, then she can disclose this by reporting $r = E_1$ or she can suppress it by reporting $r = \varphi_1$. We assume that E_1 is hard evidence that can be credibly disclosed, so if P observed φ_1, then P can only report φ_1.

Given that exculpatory evidence is perfect, a report of $r = E_1$ means that the case is dropped since the exculpatory evidence exonerates D and P can save the costs of trial preparation as well as trial itself by dropping the case. But if P does not disclose exculpatory evidence (i.e., P suppresses the observation of E_1, or P actually observed φ_1), then P makes a plea offer S_O to D at the end of period 1, which D either accepts or rejects.[25] If D accepts P's plea offer, then D is convicted and the game moves to the post-conviction stage (period 3), which will be described in detail below.

If D rejects the plea offer S_O, then period 2 proceeds. At this point, we assume that a P who

[24] For convenience, we will sometimes denote P's types as φ_1 (uninformed) and E_1 (informed); thus, γ is the probability that P is type E_1 given that D is innocent.

[25] We assume that P makes a take-it-or-leave-it plea offer. This reflects the power prosecutors have (the state enforces P's position of authority) and that P is a repeat player (against the population of defendants) and benefits from developing a reputation for not haggling. Kutateladze and Andriloro (2014) document a two-year study of the New York County (i.e., Manhattan) District Attorney's office and indicate that "The DANY adheres to the so called 'best-offer-first' approach, in which ADAs are encouraged to make the best possible offer first to save investigative resources and increase defendants' likelihood to accept the plea." (p. 134).

observed E_1 but reported $r = \varphi_1$ has a second opportunity to disclose it and/or drop the case (whether P would want to do this will depend on the anticipated penalties for suppression).[26] If the case is not dropped, then P and D expend trial preparation costs during period 2; these per-player costs are (for simplicity) assumed to be equal, and are denoted as c_2. In addition, it is now D who has an opportunity to observe a random variable denoted θ_2, which can be either $\theta_2 = E_2$, meaning that D discovered exculpatory evidence in period 2, or $\theta_2 = \varphi_2$, meaning that D did not observe exculpatory evidence in period 2. Discovery of exculpatory evidence by D during period 2 occurs with probability $\eta_2 \equiv \Pr\{\theta_2 = E_2 \mid \text{D is I}\}$.[27] If D observes E_2, then she will definitely disclose it, as the case will be dismissed immediately. If D does not observe E_2 (either because D is I but didn't discover E_2, or because D is G), then the probability of conviction at trial remains π.

If D did not observe E_2, then at the end of period 2 but before trial, an informed P makes a final decision to either disclose E_1 and drop the case, or to go to trial (below we provide conditions for the model's parameters such that an uninformed P will never voluntarily drop the case). If the case is not dropped then P and D each expend trial costs; these are (for simplicity) assumed to be equal, and denoted c_T. At trial, D is convicted with probability π, as neither P nor D has disclosed any exculpatory evidence.

We will refer to the post-conviction stage as period 3, whether it follows a plea deal or a conviction at trial.[28] During this period, we assume that an innocent D has an opportunity to observe

[26] Basically, we will allow an informed P to disclose E_1/drop the case after every informative event. For example, D's rejection of a plea offer is one such event.

[27] Since we have assumed there is no exculpatory evidence if D is G, then $\Pr\{\theta_2 = E_2 \mid \text{D is G}\} = 0$.

[28] Since we are not incorporating discounting, the start date of the post-conviction stage is not important. In any event, period 2 (the trial preparation stage) is likely to be substantially shorter than period 3 (the post-conviction stage), but we expect that both are likely to be meaningfully longer than period 1. Such likely relative lengths of the periods do not enter our analysis.

a random variable denoted θ_3, which can be either $\theta_3 = E_3$, meaning that D discovered exculpatory evidence in period 3, or $\theta_3 = \varphi_3$, meaning that D did not observe exculpatory evidence in period 3. Discovery of perfect exculpatory evidence during period 3 by D occurs with probability $\eta_3 \equiv \Pr\{\theta_3 = E_3 \mid \text{D is I}\}$.[29] If D observes E_3, then he should definitely disclose it, as it exonerates him. Following an exoneration, we assume that P will be investigated to determine whether she suppressed evidence. The penalty for verified suppression will depend on the disclosure regime that applies; we specify the regime-specific penalties in the Appendix.

Figure 1 summarizes the foregoing sequence of events and the possible changing states of information available to P and to D. In the Figure we provide the changes in information above the heavy black time line, and actions for the agents to take below the heavy black time line.

Place Figure 1 about here
----

4. Analysis of the Alternative Disclosure Regimes

We consider three disclosure regimes to which a prosecutor may be subject:

(1) no required disclosure of exculpatory evidence, denoted as N;

(2) perfectly enforced disclosure of exculpatory evidence before trial (perfect enforcement entails a high enough expected penalty to P from violating the disclosure requirement). Our shorthand for this regime is "*Brady* disclosure" and it is denoted as B;

(3) perfectly enforced full disclosure from arrest onward. This is denoted as X (for extensive disclosure).

We employ the superscripts N, B, and X as needed in the analysis. We use the letter U for

[29] This probability may appear to be small, but it is a typical trigger for exoneration and investigation. Moreover, exculpatory evidence could be found by an Innocence Project or a Conviction Integrity Unit of a District Attorney's Office.

P's payoff and the letter V for D's payoff. Subscripts are used to track where on the timeline in Figure 1 we are computing a particular payoff. In each regime, we assume that the informed P is incentivized to disclose as specified in the regime; formal statements about expected penalties that ensure this are provided and discussed in the Appendix.

4.1. Analysis of Regime N

In this subsection, P is under no obligation to disclose exculpatory evidence to D. Our purpose for doing this is to provide a point of comparison for the two alternative regimes (B and X) that require differing timing of disclosure.

4.1.1. P's Decision Problem Following Rejection and at Trial

We work backward from the last decision to be made, which is P's decision about whether to drop the case or proceed to trial (just before period 3 commences; see Figure 1). In order to have arrived at this point, the parties had to go through periods 1 and 2 without anyone having disclosed exculpatory evidence. But in that case, it is common knowledge that D will be convicted at trial with probability π (based on the original evidence provided at time zero by the police).

P's expected payoff from trial, when she reported r and observed θ_1, is denoted $U^N_T(r; \theta_1)$. As indicated in Section 3, superscript N indicates that it is P's payoff in regime N; the subscript T indicates that this is the continuation payoff starting at the point of trial (see Figure 1). For an informed P who reported φ_1, $U^N_T(\varphi_1; E_1) = \pi S_T(1 - \eta_3) - c_T$; that is, P expends c_T and wins the case with probability π, but this may be reversed later (in period 3) if D discovers exculpatory evidence. We assume that this payoff is strictly positive,[30] so an informed P who has suppressed exculpatory evidence and arrived at the trial stage will not disclose and drop the case. Since an uninformed P's

[30] The formal statement is given in Assumption A1 in the Appendix.

expected payoff from trial exceeds that of an informed P (because the uninformed P expects to face a mixture of guilty and innocent defendants, whereas the informed P knows that she faces an innocent D who may later discover exculpatory evidence), an uninformed P who has arrived at the trial stage will not drop the case. Rather, in regime N, both types of P have a credible threat to take the case to trial.

Now consider P's decision about whether to drop the case following D's rejection of a plea offer. We denote this payoff by $U^N_R(r;\ \theta_1)$, where the subscript R indicates that this is the continuation payoff starting at the point of rejection of the plea offer (i.e., at the beginning of period 2 in Figure 1). For an informed P who reported φ_1, $U^N_R(\varphi_1;\ E_1) = (1 - \eta_2)U^N_T(\varphi_1;\ E_1) - c_2 = (1 - \eta_2)[\pi S_T(1 - \eta_3) - c_T] - c_2$; that is, if D rejects an offer and P continues with the case, then P expends the trial preparation cost c_2 and only goes to trial if D does not discover exculpatory evidence during period 2. We assume that this payoff is strictly positive,[31] so an informed P who has suppressed exculpatory evidence and whose plea offer was rejected will not disclose and drop the case at this point. Again, since an uninformed P's expected payoff from continuing following a rejected plea offer exceeds that of an informed P (because the uninformed P expects to face a mixture of guilty and innocent defendants, whereas the informed P knows that she faces an innocent D, who may later discover exculpatory evidence), an uninformed P will never drop the case following a rejected plea offer. Thus, in regime N, both types of P have a credible threat to continue following rejection of the plea offer.

4.1.2 D's Decision at the Plea Bargaining Stage

Now we consider the decision problem facing D. Since D has private information about his

[31] The formal statement is given in Assumption A2 in the Appendix.

type (G or I), D thereby has some information about P's type (informed or uninformed). In particular, if D is type G then D knows that P is uninformed (because there is no exculpatory evidence); whereas if D is type I then D knows that P is informed with probability γ and uninformed with probability $1 - \gamma$. These different beliefs will prove to be important when we consider regimes with disclosure enforcement (i.e., B or X), but in regime N both types of P have a credible threat to pursue the case following a rejected plea offer.

Recall that not only is D's guilt or innocence his private information, so is his incremental disutility (loss) from conviction (δ). First consider the decision problem of type (G, δ), given a plea offer of S_O. Let A denote acceptance and R rejection (both actions by D) of P's plea offer. We denote a guilty D's disutility of accepting S_O by $V_A^N(S_O; G, \delta) = S_O + \delta$. On the other hand, a guilty D's disutility of rejecting S_O is given by $V_R^N(S_O; G, \delta) = \pi(S_T + \delta) + (c_2 + c_T)$; the trade-off here is that D faces only a probability π of being convicted, but he has to expend the costs associated with preparing for, and conducting, the trial.

Comparing these payoffs yields a threshold value of δ, denoted $\delta_G^N(S_O)$, such that a D of type (G, δ) should accept the plea offer if $\delta \leq \delta_G^N(S_O)$ and otherwise reject it. Direct computation yields:

$$\delta_G^N(S_O) = [\pi S_T + (c_2 + c_T) - S_O]/(1 - \pi). \quad (1)$$

Notice that $\delta_G^N(S_O)$ is decreasing in S_O; that is, the higher is the plea offer, the smaller is the set of guilty defendants who are willing to accept.[32] Also, $\delta_G^N(S_O)$ is increasing in π, c_2, and c_T, as increases in these reflect increased weakness of D's bargaining position: the set of guilty defendants who will accept the plea offer S_O will be larger.

Now consider the decision problem of type (I, δ), given a plea offer of S_O. We denote I's

[32] Strictly speaking, $\delta_G^N(S_O) = \max\{0, \min\{[\pi S_T + (c_2 + c_T) - S_O]/(1 - \pi), \delta^M\}\}$. We will restrict the parameter space so that all thresholds $\in (0, \delta^M)$.

disutility of accepting S_O by $V^N_A(S_O; I, \delta)$, and it is straightforward to see that $V^N_A(S_O; I, \delta) = (S_O + \delta)(1 - \eta_3)$. Although acceptance results in a conviction, an innocent D has a chance η_3 to discover exonerating evidence in the post-conviction stage. On the other hand, I's disutility of rejecting S_O is given by:

$$V^N_R(S_O; I, \delta) = (1 - \eta_2)(1 - \eta_3)\pi(S_T + \delta) + c_2 + (1 - \eta_2)c_T.$$

This payoff is explained as follows: upon rejecting S_O, a D of type I anticipates that he will expend c_2 but he will have a chance (η_2) to discover exculpatory evidence during period 2, in which event the case against him will be dismissed. If he does not discover exculpatory evidence in period 2, he will go to trial (at a cost of c_T) and may be convicted, but he will have a chance (η_3) to be exonerated post-conviction.

Comparing $V^N_A(S_O; I, \delta)$ with $V^N_R(S_O; I, \delta)$ yields a threshold value of δ, denoted $\delta^N_I(S_O)$, such that a D of type (I, δ) should accept the plea offer if $\delta \leq \delta^N_I(S_O)$ and otherwise reject it. Direct computation yields:

$$\delta^N_I(S_O) = \{(1 - \eta_3)[(1 - \eta_2)\pi S_T - S_O] + c_2 + (1 - \eta_2)c_T\}/[(1 - \eta_3)(1 - (1 - \eta_2)\pi)]. \quad (2)$$

The threshold $\delta^N_I(S_O)$ is also decreasing in S_O; that is, the higher is the plea offer, the smaller is the set of I-type defendants who are willing to accept. Similar to $\delta^N_G(S_O)$, the threshold $\delta^N_I(S_O)$ is increasing in the strength of P's case and in the cost of preparing for (and pursuing) trial.

Notice that if $\eta_2 = \eta_3 = 0$, then $\delta^N_I(S_O) = \delta^N_G(S_O)$. Moreover, as long as $\delta^N_I(S_O) > 0$,[33] then $\partial\delta^N_I(S_O)/\partial\eta_2 < 0$ and $\partial\delta^N_I(S_O)/\partial\eta_3 > 0$. That is, the set of I-types that are willing to accept the plea offer shrinks (resp., expands) as the chance of discovering exculpatory evidence in period 2, prior to trial (resp., in period 3, post-conviction) increases. Thus, in principle, it is possible to have $\delta^N_I(S_O)$

[33] We make further restrictions on parameters (see the Appendix) so as to ensure that both thresholds are interior (i.e., between 0 and δ^M); some Gs accept, and others reject, the plea offer, and similarly for some Is.

$< \delta^N_G(S_O)$ or $\delta^N_I(S_O) > \delta^N_G(S_O)$. This latter situation could arise if η_3 was relatively large and η_2 was relatively small. However, this seems unlikely as D's counsel will presumably be searching intensively during period 2 to discover exculpatory evidence in advance of trial, whereas the discovery of exonerating evidence post-conviction may be more fortuitous than the result of intense search (e.g., new forensic techniques might be developed, or the real perpetrator may confess or be identified). On the other hand, the post-conviction stage is longer than the trial preparation stage. All in all, it seems most plausible to have $\delta^N_I(S_O) < \delta^N_G(S_O)$; that is, the set of innocent Ds who are willing to accept a plea offer is smaller than the corresponding set of guilty Ds. Note, however, that our analysis does not impose this condition.

4.1.3. P's Optimal Plea Offer

The thresholds for acceptance in equations (1) and (2) do not depend on D's beliefs about P's type, since both informed and uninformed Ps have a credible threat to continue through trial following a rejection. Consequently, neither type of P could gain by mimicking the other's plea offer. Each type of P will simply make her optimal offer, and the equilibrium will be revealing.[34]

Let $U^N_O(S_O; r, \theta_1)$ denote the expected payoff (under regime N) to a P who offers S_O, after having reported r but observed θ_1. Then the expected payoff to an uninformed P is:

$$U^N_O(S_O; \varphi_1, \varphi_1) = \lambda_1\{(\delta^N_I(S_O)/\delta^M)S_O(1 - \eta_3) + (1 - \delta^N_I(S_O)/\delta^M)((1 - \eta_2)[\pi S_T(1 - \eta_3) - c_T] - c_2)\}$$
$$+ (1 - \lambda_1)\{(\delta^N_G(S_O)/\delta^M)S_O + (1 - \delta^N_G(S_O)/\delta^M)[\pi S_T - c_T - c_2]\}. \quad (3)$$

The expression in the first set of curly brackets in equation (3) is what an uninformed P would obtain if facing an I, while the expression in the second set of curly brackets in equation (3)

[34] Again, since an I knows he is innocent, he already knows that some sort of exculpatory evidence exists. However, even if he infers (from the plea offer) that P has observed exculpatory evidence and is suppressing it, this does not enable him to improve his chances of discovering E in periods 2 or 3. We discuss the implications of relaxing this assumption in subsection 6.2

is what an uninformed P would obtain if facing a G. This payoff function is a strictly concave (quadratic) function of S_O over the relevant range of possible offers;[35] hence it has a unique maximizer which is the solution to the first-order condition. The unique maximizer of equation (3) is the uninformed P's optimal plea offer, which is denoted $S_O^N(\varphi_1)$. To understand its properties, it is worth focusing for a moment on what an optimal plea offer would be if P "targeted" only the type-I or only the type-G defendants. Let $s_O^N(I)$ be the maximizer of the first term above in curly brackets; this would maximize P's payoff from only I-types. It is straightforward to show that $s_O^N(I) = (1 - \eta_2)\pi S_T$. Similarly, let $s_O^N(G)$ maximize the second term in equation (3) (that is, the second term above in curly brackets); this would maximize P's payoff from only G-types. It is straightforward to show that $s_O^N(G) = \pi S_T$. Because of the overall form of equation (3) and the fact that it is a convex combination of the two curly-bracketed terms, then the overall optimizer, $S_O^N(\varphi_1)$, lies in the interval $(s_O^N(I), s_O^N(G)) = ((1 - \eta_2)\pi S_T, \pi S_T)$. Specifically:[36]

$$S_O^N(\varphi_1) = \alpha^N s_O^N(I) + (1 - \alpha^N)s_O^N(G), \tag{4}$$

where

$$\alpha^N \equiv \lambda_1(1 - \eta_3)(1 - \pi)/\{\lambda_1(1 - \eta_3)(1 - \pi) + (1 - \lambda_1)(1 - (1 - \eta_2)\pi)\}. \tag{5}$$

Now consider the optimal plea offer for an informed P who has suppressed the observed exculpatory evidence; the relevant expected payoff is:

$$U_O^N(S_O; \varphi_1, E_1) = (\delta_I^N(S_O)/\delta^M)S_O(1 - \eta_3) + (1 - \delta_I^N(S_O)/\delta^M)((1 - \eta_2)[\pi S_T(1 - \eta_3) - c_T] - c_2). \tag{6}$$

This expression is exactly what appears inside the first set of curly brackets in equation (3); we

[35] The "relevant range" of offers are those for which the threshold functions for G and I are both strictly between 0 and δ^M. Very low offers (that are accepted for sure) or very high offers (that are rejected for sure) add "kinks" to P's payoff function. In the Appendix, we provide sufficient conditions on the parameters that guarantee that the optimal offer lies in the relevant range.

[36] See the Appendix for the explicit formula for $S_O^N(\varphi_1)$.

know the unique maximizer of this objective function, which we denote as $S_O^N(E_1)$, is $s_O^N(I) = (1 - \eta_2)\pi S_T$; that is:

$$S_O^N(E_1) = (1 - \eta_2)\pi S_T. \qquad (7)$$

The results for regime N are summarized in the following theorem.

Theorem 1. When there is no disclosure requirement (regime N):

(i) The equilibrium is revealing; an informed P makes a plea offer of $S_O^N(E_1)$, and an uninformed P makes a plea offer of $S_O^N(\varphi_1)$, where $S_O^N(\varphi_1) > S_O^N(E_1)$; see equations (4), (5), and (7) for the relevant formulas.

(ii) A D whose type is (G, δ) accepts $S_O^N(\varphi_1)$ for values of $\delta \leq \delta_G^N(S_O^N(\varphi_1))$ (see equation (1)) and otherwise rejects the offer.

(iii) A D whose type is (I, δ) accepts, if offered, $S_O^N(\varphi_1)$ for values of $\delta \leq \delta_I^N(S_O^N(\varphi_1))$ (see equation (2)) and otherwise rejects the offer. If P instead offers $S_O^N(E_1)$, then D accepts if $\delta \leq \delta_I^N(S_O^N(E_1))$ (see equation (2)) and otherwise rejects the offer. Since $S_O^N(\varphi_1) > S_O^N(E_1)$, the set of those Ds accepting the plea offer $S_O^N(\varphi_1)$ is smaller than the set of those accepting the plea offer $S_O^N(E_1)$.

(iv) P (whether informed or uninformed) never drops the case voluntarily following a rejected plea offer.

4.2. Analysis of Regime B

In this subsection, we assume that there is effective enforcement of the *Brady* rule with respect to trial. That is, P has an obligation to disclose any exculpatory evidence prior to trial and, if it is subsequently verified that she did not comply with this obligation, she will be subject to a penalty. To say that there is "effective enforcement" means that the expected penalty is sufficiently high that P prefers to disclose rather than suppress exculpatory evidence at the point of trial.[37] However, in this subsection we assume that there is no penalty for obtaining a false conviction via plea; this is consistent with how things operate in most jurisdictions and with *U.S. v. Ruiz* (2002).

[37] See Assumption A3 in the Appendix for the formal statement.

As indicated earlier, all of the relevant payoffs and equilibrium strategies will be superscripted by B (instead of N), and we refer to this as regime B.

4.2.1 P's Decision Problem Following Rejection and at Trial

In regime B, the informed P is induced to disclose and drop the case if it reaches the point of trial. Therefore, she should disclose and drop the case immediately following rejection of a plea offer (else she will incur trial preparation costs and will subsequently drop the case anyway if it reaches trial). Since an uninformed P (who observed φ_1 and reported φ_1) expects to face a mixture of innocent and guilty Ds, and is not subject to a penalty even if D later discovers exculpatory evidence, an uninformed P will never voluntarily drop the case. These observations in turn imply that in regime B there cannot be a revealing equilibrium wherein an informed P makes a different plea offer than an uninformed P.[38] For if this were to happen, then an informed P's revealing offer would always be rejected, because D (who knows that he is innocent, and who now believes that P also knows he is innocent) would anticipate that P would disclose and drop the case following rejection. The informed P would do better by deviating to mimic the uninformed P's plea offer, hoping to collect a conviction at the plea bargaining stage, knowing that she can always drop the case if D rejects the plea. Mimicry is valuable here because an uninformed P has a credible threat to continue with the case after a rejected plea offer, and the informed P can thus gain by persuading the type-I D that she is an uninformed P. Thus, in what follows, we hypothesize a pooling equilibrium wherein an uninformed P chooses her optimal plea offer, recognizing that the informed

[38] See the Appendix for a formal proof of the non-existence of a revealing equilibrium. Here we provide an intuitive discussion.

P cannot be deterred from making the same offer.[39]

4.2.2. D's Decision at the Plea Bargaining Stage

Consider the decision of a D of type G who has received a plea offer of S_O. Since D knows he is G, he also knows that P is uninformed (i.e., P observed $\theta_1 = \varphi_1$), and that P will never voluntarily drop the case if D rejects the offer. The type-G defendant's payoffs are exactly the same in regime B as in regime N: $V_A^B(S_O; G, \delta) = S_O + \delta$ and $V_R^B(S_O; G, \delta) = \pi(S_T + \delta) + (c_2 + c_T)$. The resulting threshold value, denoted $\delta_G^B(S_O)$, is also exactly the same as in regime N:

$$\delta_G^B(S_O) = \delta_G^N(S_O) = [\pi S_T + (c_2 + c_T) - S_O]/(1 - \pi). \tag{8}$$

Now consider the decision of a D of type I who has received a plea offer of S_O. Since D knows he is I, he also knows that P is informed with probability γ (and will therefore drop the case should D reject the plea offer) and uninformed with probability $1 - \gamma$ (and will therefore continue with the case should D reject the plea offer). The type-I defendant's expected payoff from accepting the plea offer S_O is the same as in regime N: $V_A^B(S_O; I, \delta) = V_A^N(S_O; I, \delta) = (S_O + \delta)(1 - \eta_3)$. However, the expected payoff from rejecting the plea offer is now: $V_R^B(S_O; I, \delta) = (1 - \gamma)V_R^N(S_O; I, \delta)$ $= (1 - \gamma)\{(1 - \eta_2)(1 - \eta_3)\pi(S_T + \delta) + c_2 + (1 - \eta_2)c_T\}$. This is because D expects that P is uninformed with probability $1 - \gamma$, in which case the remainder of the game plays out the same as in regime N, but with probability γ, P is informed and will drop the case following a rejection. Comparing these two payoffs yields a threshold value of δ, denoted $\delta_I^B(S_O)$, such that a D of type (I, δ) should accept the plea offer if $\delta \leq \delta_I^B(S_O)$ and otherwise reject it. Direct computation yields:

$$\delta_I^B(S_O) = \{(1 - \eta_3)[(1 - \gamma)(1 - \eta_2)\pi S_T - S_O] + (1 - \gamma)(c_2 + (1 - \eta_2)c_T)\}/[(1 - \eta_3)(1 - (1 - \gamma)(1 - \eta_2)\pi)]. \tag{9}$$

The threshold $\delta_I^B(S_O)$ is decreasing in S_O; that is, the higher is the plea offer, the smaller is the set of

[39] There are multiple pooling equilibria for this problem. However, in the Appendix we provide a forward induction argument that selects the equilibrium developed in the remainder of this subsection.

I-type defendants who are willing to accept. Moreover, as in regime N, as long as $\delta_I^B(S_O) > 0$, then $\partial\delta_I^B(S_O)/\partial\pi > 0$, $\partial\delta_I^B(S_O)/\partial\eta_2 < 0$, and $\partial\delta_I^B(S_O)/\partial\eta_3 > 0$. Finally, as long as $\delta_I^B(S_O) > 0$, then $\partial\delta_I^B(S_O)/\partial\gamma < 0$. That is, the set of I-type defendants who are willing to accept the plea offer is smaller, the higher is the probability that P is informed (and thus P is anticipated to drop the case following a rejection). This further implies that:

$$\delta_I^B(S_O) < \delta_I^N(S_O); \qquad (10)$$

that is, the set of Ds of type I that accept the plea offer S_O is smaller in regime B than in regime N.

In the case of $\gamma = \eta_2 = \eta_3 = 0$, then $\delta_I^B(S_O) = \delta_G^B(S_O)$; in this case, there is no fundamental distinction between Ds of type I versus G, so their thresholds are the same. Increases in γ and η_2 make the D of type I more resistant to accepting a plea (they are more willing to continue in hopes of having the case dropped or finding exculpatory evidence during trial preparation), whereas an increase in η_3 makes the D of type I more willing to accept a plea (with higher hopes of finding exculpatory evidence post-conviction). As in Subsection 4.1.2, it is possible (through a choice of values for the parameters γ, η_2 and η_3) to have either $\delta_I^B(S_O) < \delta_G^B(S_O)$ or $\delta_I^B(S_O) > \delta_G^B(S_O)$; but it seems most plausible to have $\delta_I^B(S_O) < \delta_G^B(S_O)$; that is, the set of innocent Ds who are willing to accept a plea offer is smaller than the corresponding set of guilty Ds. Again, we do not impose this inequality.

4.2.3. P's Optimal Plea Offer

Let $U_O^B(S_O; r, \theta_1)$ denote the payoff in regime B to a P who offers S_O at the beginning of period 2 (see Figure 1), having observed θ_1 and reported r. The payoff function for such an uninformed P is:

$$U_O^B(S_O; \varphi_1, \varphi_1) = \lambda_1\{(\delta_I^B(S_O)/\delta^M)S_O(1 - \eta_3) + (1 - \delta_I^B(S_O)/\delta^M)((1 - \eta_2)[\pi S_T(1 - \eta_3) - c_T] - c_2)\}$$

$$+ (1 - \lambda_1)\{(\delta^B_G(S_O)/\delta^M)S_O + (1 - \delta^B_G(S_O)/\delta^M)[\pi S_T - c_T - c_2]\}, \quad (11)$$

since an uninformed P computes her expected payoff for a plea offer of S_O using the thresholds $\delta^B_I(S_O)$ and $\delta^B_G(S_O)$.

As in regime N, this objective function is a strictly concave (quadratic) function of S_O for offers in the relevant range (see footnote 35); hence it has a unique maximizer which is the solution to the first-order condition. We will shortly provide the uninformed P's optimal plea offer, denoted $S^B_O(\varphi_1)$. However, because it is a rather complicated function of the parameters, it is again worth focusing for a moment on what an optimal plea offer would be if P "targeted" only the type-I or only the type-G defendants. Similar to the regime N analysis, let $s^B_O(I)$ be the maximizer of the first term above in curly brackets; this would maximize P's payoff from only I-types. Some algebra yields:

$$s^B_O(I) = \{(1 - \eta_3)(1 - \eta_2)(2 - \gamma)\pi S_T - \gamma(c_2 + (1 - \eta_2)c_T)\}/2(1 - \eta_3), \quad (12)$$

which is strictly less than $s^N_O(I) = (1 - \eta_2)\pi S_T$. This reflects the fact that because an uninformed P recognizes that she cannot deter mimicry by an informed P, then a D of type I will respond more skeptically to any plea offer P makes (that is, the set of type-I D's that accept will be smaller, since they anticipate a dropped case, after rejection, with probability $\gamma > 0$). Thus, if the uninformed P were to specifically "target" the D of type I, then she would make a lower offer in regime B than in regime N (because of I's greater resistance to accepting any given plea offer in regime B).

Continuing the analysis of the maximization of $U^B_O(S_O; \varphi_1, \varphi_1)$, let $s^B_O(G)$ be the maximizer of the second term in curly brackets in equation (11) above; this would maximize P's payoff from only G-types. It is straightforward to show that this offer is the same as in regime N:

$$s^B_O(G) = s^N_O(G) = \pi S_T. \quad (13)$$

Returning to the overall problem of determining the optimal plea offer, we see that the overall optimizer, $S^B_O(\varphi_1)$, lies in the interval $(s^B_O(I), s^B_O(G))$. Specifically:[40]

$$S^B_O(\varphi_1) = \alpha^B s^B_O(I) \ + (1 - \alpha^B)s^B_O(G), \tag{14}$$

where:

$$\alpha^B \equiv \lambda_1(1 - \eta_3)(1 - \pi)/\{\lambda_1(1 - \eta_3)(1 - \pi) + (1 - \lambda_1)(1 - (1- \eta_2)(1 - \gamma)\pi)\}. \tag{15}$$

Since, as discussed earlier, both types of P make the same offer in equilibrium, then $S^B_O(E_1) = S^B_O(\varphi_1)$. As long as $\delta^B_I(S^B_O(\varphi_1)) > 0$ (that is, as long as some innocent Ds accept), then the informed P receives a strictly positive payoff of $U^B_O(S^B_O(\varphi_1); \varphi_1, E_1) = (\delta^B_I(S^B_O(\varphi_1))/\delta^M)S^B_O(\varphi_1)(1 - \eta_3)$, as she simply drops the case following a rejection. If an informed P deviated at all from $S^B_O(\varphi_1)$, then a type-I defendant would infer that P is informed (i.e., in possession of exculpatory evidence) and would reject for sure, anticipating that P will subsequently drop the case because of the perfectly-enforced *Brady* rule. Finally, if an informed P chose to disclose rather than making a plea offer, then she would also obtain a payoff of zero. Thus, the fact that the *Brady* rule is effectively enforced at trial is not sufficient to induce an informed P to disclose exculpatory evidence before making a plea offer. Instead, she attempts (bluffs) to get a conviction via plea bargaining, and only discloses the exculpatory evidence and drops the case if I rejects the plea offer.

The results for regime B are summarized in the following theorem.

Theorem 2. When disclosure is required before trial (regime B):

(i) The equilibrium involves pooling; both an informed P and an uninformed P make a plea offer of $S^B_O(E_1) = S^B_O(\varphi_1)$; see equations (14) and (15) for the relevant formulas.

(ii) A D whose type is (G, δ) accepts $S^B_O(\varphi_1)$ for values of $\delta \leq \delta^B_G(S^B_O(\varphi_1))$ (see equation (8)) and otherwise rejects the offer.

(iii) A D whose type is (I, δ) accepts $S^B_O(\varphi_1)$ for values of $\delta \leq \delta^B_I(S^B_O(\varphi_1))$ (see equation (9)) and otherwise rejects the offer.

[40] See the Appendix for the explicit formula for $S^B_O(\varphi_1)$.

(iv) An uninformed P never drops the case voluntarily following a rejected plea offer; an informed P drops the case voluntarily following a rejected plea offer.

4.3. Analysis of Regime X

Next, consider a regime wherein disclosure is enforced from the point of arrest onward (extensive disclosure); the superscript X indexes the thresholds for acceptance and the plea offer. In regime X, we assume that the *Brady* rule is still enforced at the trial stage (so if a case with an informed P were to advance past the plea bargaining stage, it would be optimal for P to disclose/drop the case following a rejected plea offer[41]). Moreover, in order to induce an informed P to disclose exculpatory evidence (which yields a payoff of zero) instead of making a plea offer, there must be an expected penalty following a conviction by plea bargain that is high enough to render an informed P's expected payoff from making a plea offer negative.[42] But if informed Ps are expected to disclose, then the receipt of a plea offer must be inferred to be coming from an uninformed P, who will never voluntarily drop the case. This means that: (1) the defendants will use the same thresholds as in regime N: $\delta^X_I(S_O) = \delta^N_I(S_O)$ and $\delta^X_G(S_O) = \delta^N_I(S_O)$; (2) the uninformed P's expected payoff is the same as in regime N: $U^X_O(S_O; \varphi_1, \varphi_1) = U^N_O(S_O; \varphi_1, \varphi_1)$; and (3) the uninformed P's plea offer in regime X is the same as in the case of no enforcement (because in both of these cases, the uninformed P will never voluntarily drop the case): $S^X_O(\varphi_1) = S^N_O(\varphi_1)$.

The results for regime X are summarized in the following theorem.[43]

[41] We assume that P can disclose/drop the case at any time without penalty (or, equivalently, she is exposed to the risk of a suppression-related penalty only if she obtains a conviction). So if P tried to induce a plea bargain but failed, then she can disclose following a rejection and suffer no penalty. Otherwise, the extended disclosure rule might induce P to perversely continue the case following a rejection.

[42] See Assumption A4 in the Appendix for the formal statement.

[43] An alternative to disclosing and dropping the case before plea bargaining is for the informed P to make a revealing plea offer, which would be rejected by I, yielding a payoff to the informed P of zero. The payoffs in this equilibrium are the same for all parties as in the equilibrium with early disclosure, so we do not consider this one further.

Theorem 3. When disclosure is required from arrest onward (regime X):

(i) The equilibrium is revealing; an informed P discloses exculpatory evidence and drops the case instead of making a plea offer, and an uninformed P makes a plea offer of $S_O^X(\varphi_1) = S_O^N(\varphi_1)$; see equations (4) and (5) for the relevant formulas.

(ii) A D whose type is (G, δ) accepts $S_O^X(\varphi_1)$ for values of $\delta \leq \delta_G^N(S_O^X(\varphi_1))$ (see equation (1)) and otherwise rejects the offer.

(iii) A D whose type is (I, δ) accepts $S_O^X(\varphi_1)$ for values of $\delta \leq \delta_I^N(S_O^X(\varphi_1))$ (see equation (2)) and otherwise rejects the offer.

(iv) An uninformed P never drops the case voluntarily following a rejected plea offer.

4.4. Equilibrium Plea Offers, Thresholds, and Preferences

In this subsection, we discuss the relationships among the equilibrium plea offers (for P-types E_1 and φ_1) and among the equilibrium screening thresholds (for D-types I and G). We also present some results regarding the preferences of the parties over the three disclosure regimes.

Place Figure 2 about here

Figure 2 illustrates the relative positions of the plea offers under regime N ($S_O^N(E_1)$ and $S_O^N(\varphi_1)$), under regime B ($S_O^B(\varphi_1)$ and $S_O^B(E_1)$), and under regime X ($S_O^X(\varphi_1)$). The regime N offers are ordered so that $S_O^N(E_1) < S_O^N(\varphi_1) < \pi S_T$, reflecting: (1) regime N results in a revealing equilibrium with two offers and (2) the offer made by the uninformed P will adjust for the possibility that D is I and must therefore be less than πS_T. From Subsection 4.2 we know that regime B induces a pooling offer, so $S_O^B(\varphi_1) = S_O^B(E_1)$. Furthermore, natural parameter restrictions[44] imply that $S_O^B(\varphi_1) < S_O^N(\varphi_1)$. In the Appendix we also indicate why a sufficient condition for $S_O^B(\varphi_1) > S_O^N(E_1)$ is that the prior likelihood that D is innocent (λ_0) is sufficiently small (i.e., the police are reasonably efficient

[44] See Assumption A5 and the related discussion in the Appendix.

at making arrests, though mistakes are still made). Finally, as discussed in Subsection 4.3, if P is informed, then under regime X she will reveal the evidence and drop the case, while if she is uninformed her optimal plea offer, $S_O^X(\varphi_1)$, is the same as under regime N, $S_O^N(\varphi_1)$.

The relationships among the plea offers allow us to infer an ordering of some of the thresholds for D's decision whether to accept or reject an offer. All thresholds are decreasing in the associated plea offers: a higher plea offer results in a decrease in the marginal value (type) of δ for a D who is willing to accept the offer. Thus, $S_O^B(E_1) = S_O^B(\varphi_1) > S_O^N(E_1)$ implies that $\delta_I^B(S_O^B(\varphi_1)) < \delta_I^B(S_O^N(E_1))$, while from equation (10) we know that $\delta_I^B(S_O^N(E_1)) < \delta_I^N(S_O^N(E_1))$, yielding $\delta_I^B(S_O^B(\varphi_1)) < \delta_I^N(S_O^N(E_1))$. That is, when P is informed, I will receive a lower plea offer (and will accept it more often) in regime N than in regime B. Under the same parameter assumption as above (that λ_0 is sufficiently small[45]), we have that $\delta_I^B(S_O^B(\varphi_1)) < \delta_I^N(S_O^N(\varphi_1))$. That is, when P is uninformed, I receives a higher plea offer (but nevertheless accepts more often) in regime N than in regime B. This is because an I who is offered $S_O^B(\varphi_1)$ knows that this offer may be from an informed P, who will drop the case if the offer is rejected. Under Assumption A5 in the Appendix, and for sufficiently small λ_0, we have the following ordering results (which also rely on interiority and Assumptions A1 - A4, where applicable).

Theorem 4. Orderings of Equilibrium Offers and Thresholds.

(i) The equilibrium offers are ordered as follows:

$$S_O^N(E_1) = (1 - \eta_2)\pi S_T < S_O^B(E_1) = S_O^B(\varphi_1) < S_O^N(\varphi_1) = S_O^X(\varphi_1) < \pi S_T.$$

(ii) The equilibrium thresholds are ordered as follows:

[45] For fixed γ, the offer $S_O^B(\varphi_1)$ converges to $S_O^N(\varphi_1)$ as λ_0 goes to zero, whereas the threshold functions are independent of λ_0. We do not think that λ_0 needs to be very small. We numerically evaluated the equilibrium offers and thresholds for several parameter sets and found that these orderings of offers and thresholds held irrespective of the value of λ_0. However, due to the complexity of the formulas, we are unable to mathematically characterize the relevant range for arbitrary values of the parameters.

$$\delta_I^B(S_O^B(E_1)) = \delta_I^B(S_O^B(\varphi_1)) < \delta_I^N(S_O^N(\varphi_1)) = \delta_I^X(S_O^X(\varphi_1)) < \delta_I^N(S_O^N(E_1));$$

$$\delta_G^N(S_O^N(\varphi_1)) = \delta_G^X(S_O^X(\varphi_1)) < \delta_G^B(S_O^B(\varphi_1)).$$

Now consider the agents' preferences among regimes, evaluated at the time of arrest (that is, before P and D have an opportunity to observe exculpatory evidence). P most-prefers regime N, as it does not require any disclosure and she always has a credible threat to go to trial. A guilty defendant most-prefers regime B, because the plea offer he receives is lower in regime B than the (identical) offer he receives in regimes N and X (and the case against him will never be dropped regardless of the regime). An innocent defendant most-prefers regime X if he ends up facing an informed P (because the case will be dropped immediately in regime X, whereas P will bluff to obtain a plea agreement in regime B), and an innocent defendant most-prefers regime B if he ends up facing an uninformed P (because the plea offer he will receive is lower in regime B than in regime X). However, at the time of arrest, an innocent defendant only knows that he will face an informed P (resp., an uninformed P) with probability γ (resp., $1 - \gamma$). We are able to establish the following preference relations among the regimes (see the Appendix for details).

Theorem 5. At time zero, P most-prefers regime N; a D of type (G, δ) most-prefers regime B for all δ; and, for λ_0 sufficiently small, a D of type (I, δ) most-prefers regime X for all δ.

5. Unjust and Just Convictions

In this section we examine the three regimes (N, B, and X) and construct the probability that an innocent defendant is (unjustly) convicted of a crime. We further divide the unjust convictions into two subgroups: those wherein P was informed as to D's innocence are called "informed unjust convictions" and those wherein P was uninformed (and instead relied upon her prior and any updating that the dynamics provided) are called "uninformed unjust convictions." We also construct the probability of a "just conviction" for a guilty defendant. We then compare these measures between the three regimes. We find that the regimes X and B always dominate regime N in the sense that regimes X and B always have a (at least weakly) lower probability of informed, and uninformed, convictions of innocent defendants and a (at least weakly) higher probability of conviction of guilty defendants. However, the comparison between X and B is, surprisingly, not uniform. Regime X eliminates informed unjust convictions, but it also causes a higher plea offer to be made. This means a G will reject more often which decreases the probability of a just conviction. Moreover, if an I receives an offer under regime X, he knows that P is uninformed and will not drop the case after rejection, so I will accept more often, despite the higher offer, increasing the likelihood of uninformed unjust convictions.

5.1. Constructing Probabilities of Conviction

Consider regime N and let iuc^N denote the probability (given D = I and P is informed) of an informed unjust conviction, uuc^N denote the probability (given D = I and P is uninformed) of an uninformed unjust conviction, and jc^N denote the probability (given D = G) of a just conviction; for the other disclosure regimes we define parallel notions with superscripts B and X. Thus, for regime N, when the equilibrium offer from an informed P is $S^N_O(E_1)$, we have the probability of an informed unjust conviction of I as:

$$iuc^N = \delta^N_I(S^N_O(E_1))/\delta^M + (1 - \delta^N_I(S^N_O(E_1))/\delta^M)(1 - \eta_2)\pi,$$

where the first term accounts for (I, δ) types who accept the offer $S^N_O(E_1)$, while the second term accounts for those who reject the offer, do not discover exculpatory evidence during period 2, and are convicted at trial. Similarly, in regime N, when the equilibrium offer from an uninformed P is

$S^N_O(\varphi_1)$, we have the probability of an uninformed unjust conviction of I as:[46]

$$uuc^N = \delta^N_I(S^N_O(\varphi_1))/\delta^M + (1 - \delta^N_I(S^N_O(\varphi_1))/\delta^M)(1 - \eta_2)\pi.$$

Finally, in regime N, when the equilibrium offer from an uninformed P is $S^N_O(\varphi_1)$, we have the probability of a just conviction of G as:

$$jc^N = \delta^N_G(S^N_O(\varphi_1))/\delta^M + (1 - \delta^N_G(S^N_O(\varphi_1))/\delta^M)\pi.$$

In a similar manner, we derive the conviction probabilities for regime B (recall, this is a pooling equilibrium, so $S^B_O(E_1) = S^B_O(\varphi_1)$) as:

$$iuc^B = \delta^B_I(S^B_O(\varphi_1))/\delta^M + 0;$$

the second term is zero because an informed P drops the case after a rejection, so as to avoid being penalized for a *Brady* violation. Continuing:

$$uuc^B = \delta^B_I(S^B_O(\varphi_1))/\delta^M + (1 - \delta^B_I(S^B_O(\varphi_1))/\delta^M)(1 - \eta_2)\pi; \text{ and}$$

$$jc^B = \delta^B_G(S^B_O(\varphi_1))/\delta^M + (1 - \delta^B_G(S^B_O(\varphi_1))/\delta^M)\pi.$$

Lastly, we derive the conviction probabilities for regime X (recall that in this regime a separating equilibrium obtains wherein an informed P immediately discloses the exculpatory evidence) as:

$$iuc^X = 0;$$

$$uuc^X = \delta^X_I(S^X_O(\varphi_1))/\delta^M + (1 - \delta^X_I(S^X_O(\varphi_1))/\delta^M)(1 - \eta_2)\pi; \text{ and}$$

$$jc^X = \delta^X_G(S^X_O(\varphi_1))/\delta^M + (1 - \delta^X_G(S^X_O(\varphi_1))/\delta^M)\pi.$$

5.2 Comparisons of Regimes

First, we compare regime N and regime B. As shown in Theorem 4, $\delta^B_I(S^B_O(\varphi_1)) < \delta^N_I(S^N_O(E_1))$, and since iuc^B consists entirely of convictions due to acceptance of the informed P's plea offer, we

[46] The expressions for iuc^N and uuc^N reflect convictions. After conviction it is possible that during period 3 I will discover exculpatory evidence, and the conviction would be voided by a court. Adjusting for this possibility simply involves multiplying these expressions (and their analogs in regimes B and X below) by $(1 - \eta_3)$, which does not affect the comparisons.

know that $iuc^N > iuc^B$. Also, as shown in Theorem 4, when λ_0 is sufficiently small, $\delta_I^B(S_O^B(\varphi_1)) < \delta_I^N(S_O^N(\varphi_1))$. Both uuc^N and uuc^B can be viewed as convex combinations of the values 1 and $(1 - \eta_2)\pi$, yielding: $uuc^N > uuc^B$. Finally, by the same reasoning, $jc^N < jc^B$. That is, regime B results in a lower likelihood of informed unjust convictions, a lower likelihood of uninformed unjust convictions, and a higher likelihood of just convictions than obtain in regime N. It is straightforward to show that similar qualitative results hold for regime X versus regime N: (1) $iuc^X = 0 < iuc^N$; (2) $uuc^X = uuc^N$; and (3) $jc^X = jc^N$, so regime X, by eliminating informed unjust convictions (and being the same on other dimensions), reduces overall unjust convictions for Is, and leaves just convictions of Gs unchanged, as compared with regime N.

An important tradeoff arises in comparing B and X. While X eliminates informed unjust convictions, $uuc^X > uuc^B$ and $jc^X < jc^B$: uninformed unjust convictions are higher in regime X than in regime B and just convictions are lower in regime X compared with regime B.

We summarize the results in Section 5 in the following theorem.

Theorem 6. Comparison of Conviction Probabilities Across Regimes

(i) $iuc^N > iuc^B > iuc^X$; (ii) $uuc^N = uuc^X > uuc^B$; (iii) $jc^B > jc^N = jc^X$.

That is, a perfectly-enforced *Brady* rule generates the lowest probability of uninformed unjust conviction and the highest probability of just conviction, while the extended disclosure rule generates the lowest probability of informed unjust conviction. So the reduction in the likelihood of an informed unjust conviction that would come about by shifting from regime B to regime X comes at a cost of an increased likelihood of an uninformed unjust conviction of an innocent defendant and a reduction in the likelihood of a just conviction of a guilty defendant.

6. Discussion and Extensions

6.1 Summary

As we discussed in the Introduction, alternative evidence disclosure regimes have been implemented by legislatures and courts so as to limit abuse of power by prosecutors. The two primary regimes are disclosure of exculpatory evidence before trial and a broader regime involving disclosure of such evidence before a defendant must respond to a plea offer. To analyze the effects of different disclosure requirements, we develop a dynamic model of the disposition of a criminal case, allowing for the possibility of discovery of exculpatory evidence by prosecutors (who choose whether or not to disclose this evidence) and by defendants, as the case proceeds from arrest through plea bargaining and (possibly) trial. We characterize equilibrium behavior by prosecutors and defendants, under three different disclosure regimes. We consider: (1) a regime wherein disclosure is not required (N); (2) a second regime wherein disclosure is required before trial (B); and finally (3) a third regime wherein disclosure is required from the point of arrest onward (X). Prosecutors who have privately observed exculpatory evidence choose whether to disclose it or to suppress it and make a plea offer, and if the latter what the plea offer should be; prosecutors who have not observed exculpatory evidence simply proceed with a plea offer. When no disclosure is required, the equilibrium is revealing in the sense that an informed prosecutor makes a lower offer than one who is uninformed; no case is dropped voluntarily. When disclosure is required only prior to trial, then an informed prosecutor cannot be deterred from making the same offer as an uninformed prosecutor (so as to hide this fact), but an informed prosecutor will disclose and drop the case following a rejected plea offer. The offer in this regime is less than the uninformed offer in the no-disclosure regime. Finally, when the prosecutor is required to disclose exculpatory evidence prior to plea bargaining, then an informed prosecutor discloses and drops the case, whereas an uninformed prosecutor makes the same offer as in the no-disclosure regime (and never voluntarily drops the case). In all regimes, some innocent defendants accept the plea offer and others reject it (and similarly for guilty defendants).

We find that the *ex ante* preferences are that: 1) P prefers N; 2) G prefers B; and 3) if the likelihood of the arrest of an innocent is sufficiently small, then I prefers X. Furthermore, we find that both regimes B and X, when compared with a regime not requiring any disclosure, are (at least weakly) more likely to convict the guilty and (at least weakly) less likely to convict both types of the innocent defendant. However, when we compare likelihoods of conviction between B and X, we find that disclosure regime B leads to a higher likelihood of informed unjust conviction of innocent defendants than regime X, but a lower likelihood of uninformed unjust conviction of innocent defendants and a higher likelihood of just conviction of guilty defendants.

6.2 Robustness Issues and Possible Extensions

In order to keep the model as simple as possible, we have made several assumptions. For instance, we assumed that P and D have the same costs of preparing for, and conducting, a trial. We have also assumed that their random disutility δ is drawn from the same distribution. We have abstracted from discounting, and from the fact that an exonerated D serves a partial sentence. All of these could be incorporated into the model but the added insight would be minimal and the complication substantial.

We have modeled regime N as allowing P to act with impunity. In particular, we assumed that the likelihood (η_2) that I discovers exculpatory evidence in period 2 is independent of his inference about whether P is informed or uninformed. One might conjecture that if the informed P makes a revealing plea offer, then I infers that P is informed and will be more likely to discover E_2. However, in practice a P that suppresses evidence may effectively hide it or even destroy it, so that I may in fact be less likely to discover E_2 if P has already done so. However, for the sake of argument, let us suppose that I would discover E_2 for sure if the informed P makes a revealing plea offer. Then the informed P would simply pool with the uninformed P at $S^N_O(\varphi_1)$, but would never drop the case following a rejection (so I would continue to use the rejection function δ^N_I and $S^N_O(\varphi_1)$ would be unchanged). Thus $iuc^N = \delta^N_I(S^N_O(\varphi_1))/\delta^M + (1 - \delta^N_I(S^N_O(\varphi_1))/\delta^M)(1 - \eta_2)\pi$. Since we have previously found that $\delta^N_I(S^N_O(\varphi_1))/\delta^M > \delta^B_I(S^B_O(\varphi_1))/\delta^M$, it is still true that $iuc^N > iuc^B > iuc^X$. The other comparisons of conviction rates are also unaffected.

A more challenging extension (which seems capable of affecting at least some comparisons) would be to include imperfect exculpatory evidence. Imperfection could be modeled in various ways; for instance, exculpatory evidence might simply adjust the probability of conviction as opposed to completely exonerating an innocent defendant. Moreover, exculpatory evidence might be found even for a guilty defendant. Incorporating imperfect exculpatory evidence seems like a formidable extension due to the additional Bayesian updating for P and the fact that G’s acceptance rule would also be regime-dependent (currently, G’s acceptance rule does not vary with the regime). There are several other possible extensions to the basic model, such as endogenous investigation decisions, and heterogeneous prosecutors who have some concern for causing false convictions. Prosecutors with moral concerns about convicting an innocent defendant by suppressing evidence were explored in Daughety and Reinganum (2018), but extension to the dynamic setting with plea bargaining and consideration of alternative disclosure regimes is desirable. Finally, the initial selection of defendants by the police (i.e., arrest) influences the likelihoods of unjust convictions, so a reduction in the likelihood that an arrested person is innocent also enhances the integrity of the justice system and provides another possible lever for reducing unjust convictions.

6.3 Welfare Issues

We provided some preference results in Theorem 5. In particular, if λ_0 is sufficiently small then, at the time of arrest, regime X is most-preferred by an innocent D and least-preferred by a guilty D. However, the analysis thus far does not encompass enough welfare-related aspects to enable us to select among the disclosure regimes. This social choice comes down (partly) to the classic weighting between convicting innocents versus letting guilty perpetrators go free.[47] But the tradeoff is now compounded by the externality between the two subgroups of innocent defendants. If the probability that the prosecutor is informed (γ) is relatively low, then many more innocent defendants are at risk of unjust convictions by an uninformed prosecutor than by an informed prosecutor. However, an informed unjust conviction (because it comes about due to malfeasance by the prosecutor) can be viewed as an offense committed by persons in authority against a vulnerable individual, which undermines the integrity of the judicial system. Thus, if informed unjust convictions are considered substantially more socially undesirable than uninformed unjust convictions, then society could very well prefer the extensive-disclosure regime X to the one we view as analogous to the status quo (*Brady*) regime B (of course, provided these requirements are enforced).

However, the payoffs of criminal defendants (whether guilty or innocent) encompass only a part of what should be included in a social welfare function. For instance, a full social welfare function would incorporate an individual's decision to commit a crime and the harm thus imposed

[47] "It is better that ten guilty persons escape than that one innocent suffer" (Sir William Blackstone, *Commentaries on the Laws of England*, Clarendon Press, 1765).

on others, as well as the costs of maintaining the criminal justice system (including incarceration costs and the costs of investigation, which are likely to vary with the disclosure regime). Whereas it is not always possible to eliminate error (that is, some uninformed unjust convictions are inevitable), a system that tolerates informed unjust convictions undermines belief in the criminal justice system, which might affect the willingness of victims and witnesses to provide evidence. This is an important externality that is not captured by the payoffs of the individual agents in the model. The formulation of an appropriate welfare function is left for future research.

References

Baker, Scott, and Claudio Mezzetti. 2001. "Prosecutorial Resources, Plea Bargaining, and the Decision to Go to Trial," 17 *Journal of Law, Economics, and Organization* 149-167.

Bandoyopadhyay, S., and Bryan C. McCannon. 2014. "The Effect of the Election of Prosecutors on Criminal Trials," 161 *Public Choice* 141-156.

Bebchuk, Lucian A. 1988. "Suing Solely to Extract a Settlement Offer," 17 Journal of Legal Studies 437-450.

Bjerk, David. 2007. "Guilt Shall not Escape of Innocence Suffer: The Limits of Plea Bargaining When Defendant Guilt is Uncertain," 9 *American Law and Economics Revie*w 305-329.

Boylan, Richard T., and Cheryl X. Long. 2005. "Salaries, Plea Rates, and the Career Objectives of Federal Prosecutors," 48 *Journal of Law and Economics* 627-652.

Boylan, Richard T. 2005. "What do Prosecutors Maximize? Evidence from the Careers of U.S. Attorneys," 7 *American Law and Economics Review* 379-402.

Daughety, Andrew F. and Jennifer F. Reinganum. 2008. "Communicating Quality: A Unified Model of Disclosure and Signalling," 39 *RAND Journal of Economics* 973-989.

Daughety, Andrew F. and Jennifer F. Reinganum. 2016. "Informal Sanctions on Prosecutors and Defendants and the Disposition of Criminal Cases," 32 *Journal of Law, Economics, and Organizations* 359-394.

Daughety, Andrew F. and Jennifer F. Reinganum. 2018. "Evidence Suppression by Prosecutors: Violations of the *Brady* Rule," 34 *Journal of Law, Economics, and Organizations* 475-510.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice," 48 *Journal of Economic Literature* 935-963.

Dye, Ronald. 2017. "Optimal Disclosure Decisions When There are Penalties for Nondisclosure*,"* *48 RAND Journal of Economics* 704-732.

Farmer, Amy, and Paul Pecorino. 2007. "Negative Expected Value Suits in a Signaling Model," 74 *Southern Economic Journal* 434-447.

Franzoni, Luigi A. 1999. "Negotiated Enforcement and Credible Deterrence," 109 *The Economic Journal* 509-535.

Garoupa, Nuno, and Matteo Rizzolli. 2011. "The Brady Rule May Hurt the Innocent," 13 *American Law and Economics Review* 168-200.

Gershman, Bennett L. 2007. "Litigating *Brady v. Maryland*: Games Prosecutors Play," 57 *Case Western Reserve Law Review* 531-566.

Gershman, Bennett L. 2015. *Prosecutorial Misconduct*, 2nd Ed. Thompson Reuters.

Glaeser, Edward L., Daniel P. Kessler, and Anne M. Piehl. 2000. "What do Prosecutors Maximize? An Analysis of the Federalization of Drug Crimes," 2 *American Law and Economics Review* 259-290.

Grossman, Gene M., and Michael L. Katz. 1983. "Plea Bargaining and Social Welfare," 73 *The American Economic Review* 749-757.

Grossman, Sanford. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality," 24 *Journal of Law and Economics* 461-483.

Grunwald, Ben. 2017. "The Fragile Promise of Open-File Discovery," 49 *Connecticut Law Review* 771-836.

Hooper, Laura L., Jennifer E. Marsh, and Brian Yeh. 2004. Treatment of *Brady v. Maryland* Material in United States District and State Courts' Rules, Orders, and Policies," Report to the

Advisory Committee on Criminal Rules of the Judicial Conference of the United States, Federal Judicial Center.

Jovanovic, Boyan. 1982. "Truthful Disclosure of Information," 13 *Bell Journal of Economics* 36-44.

Katz, Avery. 1990. "The Effect of Frivolous Lawsuits on the Settlement of Litigation," 10 *International Review of Law and Economics* 3-27.

Keenan, David, Deborah Jane Cooper, David Lebowitz, and Tamar Lerer. 2011. "The Myth of Prosecutorial Accountability after *Connick v. Thompson*: Why Existing Professional Responsibility Measures Cannot Protect Against Prosecutorial Misconduct," 121 *The Yale Law Journal Online*, 203-265.

Kozinski, Alex. 2015. "Criminal Law 2.0," Preface to the *Annual Review of Criminal Procedure*, 44 *Georgetown Law Journal* iii-xliv.

Kutateladze, Besiki L. and Nancy R. Andiloro. 2014. "Prosecution and Racial Justice in New York County," Technical Report, Prosecution and Racial Justice Program, Vera Institute of Justice.

Landes, William M. 1971. "An Economic Analysis of the Courts," 14 *Journal of Law and Economics* 61-108.

Matthews, Steven and Andrew Postlewaite. 1985. "Quality Testing and Disclosure," 16 *RAND Journal of Economics* 328-340.

McCannon, Bryan C. 2013. "Prosecutor Elections, Mistakes, and Appeals," 10 *Journal of Empirical Legal Studies* 696-714.

Milgrom, Paul. 1981. "Good News and Bad News: Representation Theorems and Applications," 12 *Bell Journal of Economics* 380-391.

Mosteller, Robert P. 2008. "Exculpatory Evidence, Ethics, and the Road to the Disbarment of Mike Nifong: The Critical Importance of Full Open-File Discovery," 15 *George Mason Law Review* 257-318.

Mungan, Murat C. and Jonathan Klick. 2016. "Reducing False Guilty Pleas and Wrongful Convictions through Exoneree Compensation," 59 *Journal of Law and Economics* 173-189.

Nalebuff, Barry. 1987. "Credible Pretrial Negotiation, 18 *RAND Journal of Economics* 198-210.

Reinganum, Jennifer F. 1988. "Plea Bargaining and Prosecutorial Discretion," 78 *American Economic Review* 713-728.

Schwartz, Warren F. and Abraham L. Wickelgren. 2009. "Credible Discovery, Settlement, and Negative Expected Value Suits," 40 *RAND Journal of Economics* 636-657.

Shavell, Steven. 1994. "Acquisition and Disclosure of Information Prior to Sale," 25 *RAND Journal of Economics* 20-36.

Turner, Jenia I. and Allison D. Redlich. 2016. "Two Models of Pre-Plea Discovery in Criminal Cases: An Empirical Comparison," 73 *Washington and Lee Law Review* 285-408.

Brady v. Maryland, 373 U.S. 83 (1963).

Connick v. Thompson, 563 U.S. 51 (2011).

U.S. v. Higgs, 713 F.2d 39 (1983).

U.S. v. Ruiz 536 U.S. 622 (2002).

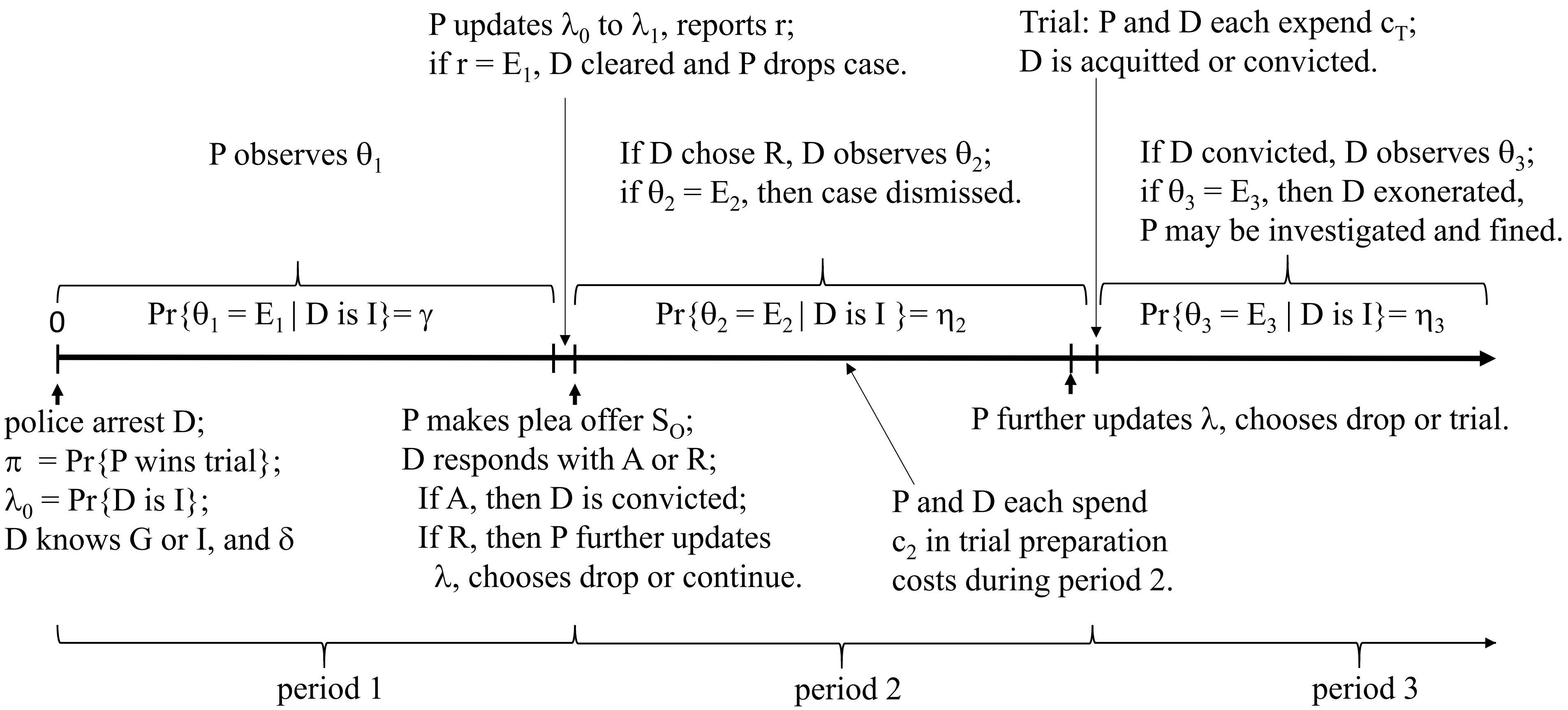
P updates λ_0 to λ_1, reports r;
if r = E_1, D cleared and P drops case.
Trial: P and D each expend c_T;
D is acquitted or convicted.
P observes θ_1
If D chose R, D observes θ_2;
if $\theta_2 = E_2$, then case dismissed.
If D convicted, D observes θ_3;
if $\theta_3 = E_3$, then D exonerated,
P may be investigated and fined.
0
$\Pr\{\theta_1 = E_1 \mid D \text{ is } I\} = \gamma$
$\Pr\{\theta_2 = E_2 \mid D \text{ is } I\} = \eta_2$
$\Pr\{\theta_3 = E_3 \mid D \text{ is } I\} = \eta_3$
police arrest D;
$\pi = \Pr\{P \text{ wins trial}\}$;
$\lambda_0 = \Pr\{D \text{ is } I\}$;
D knows G or I, and δ
P makes plea offer S_O;
D responds with A or R;
If A, then D is convicted;
If R, then P further updates
λ, chooses drop or continue.
P and D each spend
c_2 in trial preparation
costs during period 2.
P further updates λ, chooses drop or trial.
period 1
period 2
period 3

Figure 1: Sequence of Events and Information Structure

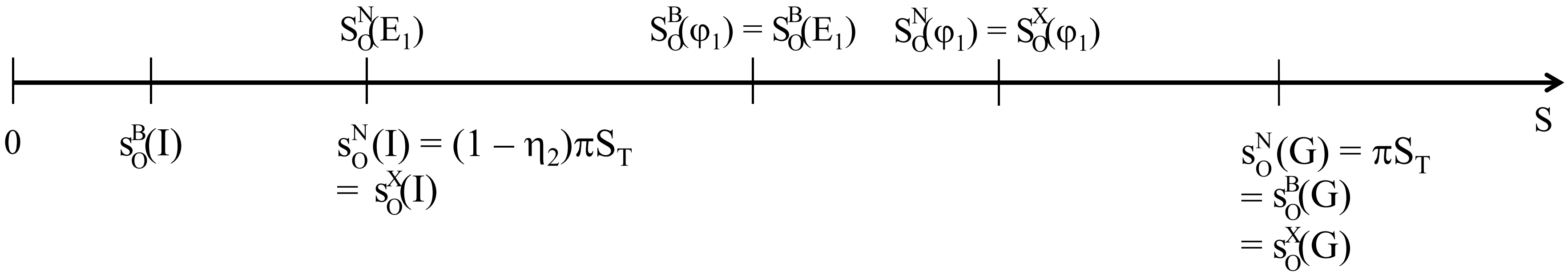
$S_O^N(E_1)$
$S_O^B(\varphi_1) = S_O^B(E_1)$
$S_O^N(\varphi_1) = S_O^X(\varphi_1)$
0
$s_O^B(I)$
$s_O^N(I) = (1 - \eta_2)\pi S_T$
$= s_O^X(I)$
$s_O^N(G) = \pi S_T$
$= s_O^B(G)$
$= s_O^X(G)$
S

Figure 2: Equilibrium Plea Offers

Appendix

Assumptions characterizing effective enforcement.

Assumptions 1 and 2 below apply in all regimes. They are sufficient to ensure that an uninformed P (who faces no penalties, and expects to face a mixture of innocent and guilty defendants at trial) will never drop the case voluntarily.

Assumption A1. $U_T^N(\varphi_1; E_1) = \pi S_T(1 - \eta_3) - c_T > 0$.

Assumption A2. $U_R^N(\varphi_1; E_1) = (1 - \eta_2)[\pi S_T(1 - \eta_3) - c_T] - c_2 > 0$.

Regime N: Assumptions A1 and A2 ensure that an informed P will never voluntarily disclose and drop the case just prior to trial (Assumption A1) and just after rejection of the plea offer (Assumption A2). *A fortiori*, an uninformed P will never drop the case at these decision points.

Regime B: Let k denote the penalty imposed on P if an investigation (following a conviction at trial) detects suppression of exculpatory evidence (that is, a "*Brady* violation"). Investigation occurs with probability η_3, since η_3 is the probability that an I discovers the exculpatory evidence post-conviction; we assume that the investigation does not generate false positives. Assumption 3 below ensures that an informed P will comply with the disclosure requirement if the case advances to the point of trial. Moreover, since an informed P would drop the case at the point of trial, then she should drop it immediately following a rejected plea offer (so as to save c_2). An uninformed P, facing no penalty, will never drop the case.

Assumption A3. $U_T^B(\varphi_1; E_1) = \pi(S_T(1 - \eta_3) - \eta_3 k) - c_T < 0$.

Regime X: We maintain Assumption 3 so that if a case with an informed P were to advance past the plea bargaining stage, it would be optimal for her to disclose/drop the case following a rejected plea offer. Moreover, in order to induce an informed P to disclose exculpatory evidence (which yields a payoff of zero) instead of making a plea offer, there must be an expected penalty following a conviction by plea bargain that is high enough to render an informed P's expected payoff from making a plea offer negative. If the informed P were going to make a plea offer (instead of disclosing), then she would have to mimic the uninformed P. Let K denote the penalty imposed on P if an investigation (following a plea-bargained conviction and discovery by D of E in period 3) later detects suppression of exculpatory evidence. The expected payoff from this mimicry is: $(\delta_I^X(S_O^X(\varphi_1))/\delta^M)[S_O^X(\varphi_1)(1 - \eta_3) - \eta_3 K]$. The following assumption induces compliance by the informed P with the extensive disclosure requirement. Again, facing no penalties, the uninformed P will never drop the case.

Assumption A4. $S_O^X(\varphi_1)(1 - \eta_3) - \eta_3 K < 0$.

Discussion regarding interiority of the thresholds.

The equilibrium plea offer in any regime will never be lower than $s_O^B(I) = \{(1 - \eta_3)(1 - \eta_2)(2 - \gamma)\pi S_T - \gamma(c_2 + (1 - \eta_2)c_T)\}/2(1 - \eta_3)$ or higher than $s_O^B(G) = s_O^N(G) = \pi S_T$, as long as the thresholds $\delta_I^B(S_O) < \delta_I^N(S_O)$ and $\delta_G^B(S_O) = \delta_G^N(S_O)$ are interior (that is, they belong to $(0, \delta^M)$) when evaluated at $s_O^B(I)$ and πS_T. Since the threshold functions are all decreasing in S_O, we want them to be greater than zero when evaluated at $S_O = \pi S_T$ and less than δ^M when evaluated at $S_O = s_O^B(I)$.

First consider the requirement that $\delta_I^B(\pi S_T)$, $\delta_I^N(\pi S_T)$ and $\delta_G^N(\pi S_T)$ should all be greater than zero. The expression $\delta_G^N(\pi S_T) = (c_2 + c_T)/(1 - \pi)$ already exceeds zero. Since $\delta_I^B(\pi S_T) < \delta_I^N(\pi S_T)$, we only need to ensure that:

$$\delta_I^B(\pi S_T) > 0. \tag{A1}$$

(A1) is satisfied if and only if: $c_2 + (1 - \eta_2)c_T > (1 - \eta_3)\pi S_T[1 - (1 - \gamma)(1 - \eta_2)]/(1 - \gamma)$. But Assumptions 1 and 2 constrain the costs to be "not too high," so we must determine whether condition (A1) is consistent with Assumptions 1 and 2. Recall that Assumption 2 implies Assumption 1, so the tightest constraint comes from Assumption 2: $c_2 + (1 - \eta_2)c_T < (1 - \eta_2)(1 - \eta_3)\pi S_T$. Thus, Assumption 2 and condition (A1) can both hold as long as $(1 - \gamma)(1 - \eta_2) > 0.5$, which is plausible.

Next, consider the requirement that $\delta_I^B(s_O^B(I))$, $\delta_I^N(s_O^B(I))$ and $\delta_G^N(s_O^B(I))$ should all be less than δ^M. Since $\delta_I^B(s_O^B(I)) < \delta_I^N(s_O^B(I))$, we need only ensure that:

$$\max\{\delta_I^N(s_O^B(I)), \delta_G^N(s_O^B(I))\} < \delta^M. \tag{A2}$$

As noted in the text, $\delta_I^N(S_O) < \delta_G^N(S_O)$ if η_3 is sufficiently small, but the reverse can occur, at least in principle. Thus, in order to ensure that all thresholds fall below δ^M (when the threshold functions are evaluated $s_O^B(I)$), it is sufficient to assume that $\delta^M > \max\{\delta_I^N(s_O^B(I)), \delta_G^N(s_O^B(I))\}$. This can be done without affecting any other parameter conditions, as δ^M is otherwise unconstrained.

Note that these are sufficient conditions for all thresholds to be interior, when evaluated at upper and lower bounds for the plea offer. However, the equilibrium plea offer will never take on either of these boundary values; rather, it will be a convex combination of $s_O^B(I)$ (or $s_O^N(I)$) and $s_O^B(G) = s_O^N(G) = \pi S_T$. The restriction that $\delta^M > \max\{\delta_I^N(s_O^B(I)), \delta_G^N(s_O^B(I))\}$ is significantly stronger than necessary since, given that λ_1 is likely to be relatively small, the convex combinations $S_O^N(\varphi_1)$ and $S_O^B(\varphi_1)$ will be heavily-weighted toward $s_O^B(G) = s_O^N(G) = \pi S_T$. On the other hand, the condition $\delta_I^B(\pi S_T) > 0$ is close to being necessary for our equilibrium thresholds to be interior.

Discussion of why there is always a pooling equilibrium at $S_O^B(\varphi_1)$, and why there cannot be a separating equilibrium, in regime B.

In the text we have argued that there is a pooling equilibrium at $S_O^B(\varphi_1)$ because the informed

P cannot be deterred from mimicking the uninformed P. We have assumed that $\delta_I^B(\pi S_T) > 0$ so that, in the pooling equilibrium described in the text, there will be some values of δ for which an I will accept the pooling offer for any $S_O < \pi S_T$. The informed P will not want to deviate from the pooling offer to a different offer, as that would reveal her to be informed, leading I to reject the plea offer for all values of δ, which yields a strictly lower payoff (of zero). Moreover, since an informed P faces no cost of making a plea offer (she can always drop the case without penalty following a rejection), an informed P strictly prefers to bluff by making a plea offer and then dropping the case if the offer is rejected as compared to disclosing and dropping the case without making a plea offer.

However, there are sufficiently high values of S_O that would result in I rejecting the plea offer for all δ, which would drive the informed P's mimicry payoff to zero. But an informed P is still willing to mimic such an offer, because her payoff is zero whether she mimics and makes the plea offer (which is rejected for sure) or deviates to some other offer (which is rejected for sure as she is inferred to be informed) or discloses and drops the case without making a plea offer. Since the informed P cannot be deterred from mimicking the uninformed P, the uninformed P's best pooling offer is $S_O^B(\varphi_1)$ as described in the text, and a pooling equilibrium always exists.

We now argue that there cannot be a separating equilibrium. How might an uninformed P try to distinguish herself? She would have to make a plea offer that is high enough to drive the informed P's payoff to zero, so that disclosure by the informed P would be a best response (even though mimicry is still always a best response). To see that there cannot be such an equilibrium, first define two critical plea offers. Let $S_1 > \pi S_T$ be defined by $\delta_I^B(S_1) = 0$; that is, $S_1 = \{(1 - \eta_3)(1 - \gamma)(1 - \eta_2)\pi S_T + (1 - \gamma)(c_2 + (1 - \eta_2)c_T)\}/(1 - \eta_3)$. Offers at this level or above would be rejected by I for all δ, based on the belief that the offer came from an informed P with probability γ and from an uninformed P with probability $1 - \gamma$. Define $S_2 > S_1$ by $\delta_I^N(S_2) = 0$, so
$S_2 = \{(1 - \eta_3)(1 - \eta_2)\pi S_T + (c_2 + (1 - \eta_2)c_T)\}/(1 - \eta_3)$. Offers at this level or above would be rejected by I for all δ, based on the belief that the offer came from an uninformed P for sure.

We know that in order to drive the informed firm's mimicry profits to zero, the offer must be at least S_1. So could there be a separating equilibrium wherein an uninformed P makes a plea offer $S_O \in [S_1, S_2)$, and an informed P discloses and drops the case (or makes some other revealing offer)? If so, then when I receives the plea offer S_O, he believes that it is coming from an uninformed P for sure, in which case he will employ the threshold $\delta_I^N(S_O) > 0$ for all $S_O \in [S_1, S_2)$, so I will accept this offer for some values of δ. But then the informed P would not be willing to disclose (or choose another offer that reveals her type), as she can now make a positive payoff by mimicking the plea offer S_O. So there cannot be a separating equilibrium wherein an uninformed P makes a plea offer $S_O \in [S_1, S_2)$, and the informed P discloses and drops the case (or makes some other revealing offer).

Finally, could there be a separating equilibrium wherein an uninformed P makes a plea offer $S_O \geq S_2$, and an informed P discloses and drops the case (or makes some other revealing offer)? If so, then when I receives the plea offer S_O, he believes that it is coming from an uninformed P for sure, in which case he will employ the threshold $\delta_I^N(S_O) = 0$ for all $S_O \geq S_2$. In this case, the informed P's payoff from mimicry is the same as her payoff from disclosing and dropping the case (or making some other revealing offer); it is zero in all three cases, so the informed P is willing to disclose or make a revealing offer. However, the uninformed P could do better by deviating to the plea offer $S = \pi S_T$, even if by doing so she was inferred (by I) to be an informed P. To see why, note that at the putative separating offer, the uninformed P is making no plea agreements with I, and is making a plea offer that exceeds the one that maximizes her payoff from G (which is $S = \pi S_T$). By deviating to $S = \pi S_T$, she will continue to make no plea agreements with I (assuming that I believes this offer is coming from an informed P, he will reject it as he anticipates the informed P will subsequently drop the case), but now she will make the maximum possible payoff from G. Thus, there cannot be a separating equilibrium wherein an uninformed P makes a plea offer $S_O \geq S_2$, and an informed P discloses and drops the case (or makes some other revealing offer).

Selection among pooling equilibria in regime B

In regime B, we have focused on the pooling equilibrium at $S^B_O(\varphi_1)$. However, there are other pooling equilibria, which are supported by beliefs that any deviation offer is coming from an informed P (which leads to a rejection by I). Recall the definition of $U^B_O(S_O; \varphi_1, \varphi_1)$ from equation (11) and the definition of $U^B_O(S_O; \varphi_1, E_1) = (\delta^B_I(S_O))/\delta^M)S_O(1 - \eta_3)$. Consider a possible pooling offer at $S_O = \bar{S}$; in order for the offer $S_O = \bar{S}$ to be a pooling equilibrium, it must be that: (1) an uninformed P prefers to offer $\bar{S}$ rather than deviating to any other offer. That is:

$$U^B_O(\bar{S}; \varphi_1, \varphi_1) \geq \lambda_1\{(1 - \eta_2)[\pi S_T(1 - \eta_3) - c_T] - c_2\}$$
$$+ (1 - \lambda_1)\{(\delta^B_G(\pi S_T)/\delta^M)\pi S_T + (1 - \delta^B_G(\pi S_T)/\delta^M)[\pi S_T - c_T - c_2]\}.$$

The right-hand-side is the uninformed P's payoff from her best deviation offer, which is $s^B_O(G) = \pi S_T$ (since I rejects any deviation offer, P's best deviation offer is the one that maximizes her payoff from a guilty defendant). In addition, it must be that: (2) an informed P prefers to offer $\bar{S}$ rather than deviating to any other offer. This simply requires that $U^B_O(\bar{S}; \varphi_1, E_1) \geq 0$, since the informed P's payoff from any deviation offer is 0 (because I rejects the offer and the informed P then drops the case). These inequalities admit a continuum of pooling equilibria; for instance, $\bar{S} \in [S^B_O(\varphi_1) - \varepsilon, S^B_O(\varphi_1) + \varepsilon]$, for some $\varepsilon > 0$.

However, among the pooling equilibria, we select the one that the uninformed P most prefers. This is because it is common knowledge to all players that the informed P cannot be deterred from mimicking the uninformed P's plea offer. Thus, the informed P should be attempting to match the uninformed P; the informed P's best guess about what the uninformed P will do is that the uninformed P will choose her most-preferred pooling offer. Likewise, the D of type I should have a conjecture about the common plea offer that will be offered by both types of P. I's best guess about that offer is the uninformed P's most-preferred pooling offer. A forward-induction argument also selects the offer $S^B_O(\varphi_1)$. This is because the uninformed P could make the following speech (and the informed P would follow suit): "We all know the equilibrium will be a pooling one; so when you observe the offer $S^B_O(\varphi_1)$ – which is my most-preferred pooling equilibrium offer – you should at least include me – the uninformed P – in the set of those you believe would make this offer." If the innocent defendant infers that the offer $S^B_O(\varphi_1)$ is coming from both types of P, then this belief allows the uninformed P to obtain her most-preferred pooling equilibrium.

Explicit formulas for equilibrium plea offers

Define the expression:

$$S_O(\varphi_1; Z) \equiv \{\pi S_T[\lambda_1(1-\eta_3)(1-\eta_2)(2-Z)(1-\pi) + 2(1-\lambda_1)(1-(1-\eta_2)(1-Z)\pi)] - \lambda_1 Z(c_2 + (1-\eta_2)c_T)(1-\pi)\}$$

$$/\{2\lambda_1(1-\eta_3)(1-\pi) + 2(1-\lambda_1)(1 - (1-\eta_2)(1 - Z)\pi)\}. \qquad \text{(A3)}$$

Then $S_O^N(\varphi_1) = S_O(\varphi_1; 0)$ and $S_O^B(\varphi_1) = S_O(\varphi_1; \gamma)$.

Orderings of equilibrium offers and thresholds

First, consider what is needed to establish that $S_O^B(\varphi_1) = S_O(\varphi_1; \gamma) < S_O^N(\varphi_1) = S_O(\varphi_1; 0)$. We treat the variable Z in $S_O(\varphi_1; Z)$ as a continuous (rather than discrete) variable, and differentiate the expression in (A3). After a great deal of algebra, it can be shown that:

sgn $\{\partial S_O(\varphi_1; Z)/\partial Z\}$ = sgn $\{Y\}$, where

$$Y \equiv \pi S_T(1 - \eta_3)(1 - \eta_2)[- (1 - \lambda_1)(1 - \pi - \eta_2\pi) - \lambda_1(1 - \eta_3)(1 - \pi)]$$

$$- (c_2 + (1 - \eta_2)c_T)[(1 - \lambda_1)(1 - \pi + \eta_2\pi) + \lambda_1(1 - \eta_3)(1 - \pi)]. \qquad \text{(A4)}$$

The expression Y is independent of Z, so either $\partial S_O(\varphi_1; Z)/\partial Z < 0$ (that is, $S_O^B(\varphi_1) < S_O^N(\varphi_1)$) for all Z , or $\partial S_O(\varphi_1; Z)/\partial Z > 0$ (that is, $S_O^B(\varphi_1) > S_O^N(\varphi_1)$) for all Z.

In particular, a <u>necessary and sufficient condition</u> for $S_O^B(\varphi_1) < S_O^N(\varphi_1)$ is $Y < 0$. A <u>sufficient condition</u> is $[- (1 - \lambda_1)(1- \pi - \eta_2\pi) - \lambda_1(1 - \eta_3)(1 - \pi)] \leq 0$; a stronger (but simpler) <u>sufficient condition</u> is $(1- \pi - \eta_2\pi) \geq 0$. In order to sustain the opposite result, all of these conditions would need to fail. However, it is plausible to entertain values of η_2 that are quite low, but this would be ruled out if the expression $(1- \pi - \eta_2\pi)$ had to be negative <u>and</u> sufficiently large in magnitude as to make $Y > 0$. Thus we continue under the parameter restriction $Y < 0$, so that $S_O^B(\varphi_1) < S_O^N(\varphi_1)$.

<u>Assumption A5</u>. The expression Y given in equation (A4) above is negative.

Next, consider what is needed to establish that $S_O^B(\varphi_1) > S_O^N(E_1) = (1 - \eta_2)\pi S_T$. Because the latter expression is the plea offer an informed P would make if she were "targeting" innocent Ds, this is strictly less than $S_O^N(\varphi_1)$. However, it is unclear whether $S_O^N(E_1) < S_O^B(\varphi_1)$ for arbitrary γ. To find a sufficient condition for this inequality to hold, recall that $\lambda_1 = \lambda_0(1 - \gamma)/[1 - \lambda_0 + \lambda_0(1 - \gamma)]$. Thus (for fixed γ), λ_1 converges to zero as λ_0 approaches zero. Moreover, from equation (A3) we see that (again, for fixed γ) the function $S_O^B(\varphi_1) = S_O(\varphi_1; \gamma)$ converges to $S_O^N(\varphi_1)$ as λ_1 goes to zero. Thus we can conclude that $S_O^N(E_1) < S_O^B(\varphi_1)$ for sufficiently small λ_0.

Finally, consider what is needed to establish that $\delta_I^B(S_O^B(\varphi_1)) < \delta_I^N(S_O^N(\varphi_1))$. In moving from regime N to regime B, there is both a direct effect on the threshold for acceptance, and an indirect effect through the change in the equilibrium plea offer. The direct effect is that I is more willing to reject any given plea offer in regime B than in regime N; that is, $\delta_I^B(S_O) < \delta_I^N(S_O)$. The indirect effect is that (under Assumption A1) the plea offer is lower in regime B (that is, $S_O^B(\varphi_1) < S_O^N(\varphi_1)$), and I is more willing to accept a lower offer. We are unable to prove, in general, whether the direct effect always outweighs the indirect effect. However, since the equilibrium plea offers $S_O^B(\varphi_1)$ and $S_O^N(\varphi_1)$ are equal in the limit as λ_0 goes to zero, whereas the threshold functions $\delta_I^B(S_O)$ and $\delta_I^N(S_O)$ are unaffected by λ_0, we can conclude that for λ_0 sufficiently small, the inequality $\delta_I^B(S_O^B(\varphi_1)) < \delta_I^N(S_O^N(\varphi_1))$ will hold.

Proof of Theorem 5. The reason for P's preference is obvious; whatever P discovers, in regime N she is able to tailor her offer to her information and she always has a credible threat of trial should her offer be rejected. Now consider the equilibrium payoff to type (G, δ) under the various regimes. Let $\psi_G(\delta) \equiv \pi(S_T + \delta) + (c_2 + c_T)$ denote this type's expected payoff from rejecting the plea offer (this is the same across all regimes). Then (G, δ)'s equilibrium payoff for regime R can be written as follows: $\min\{S_O^R(\varphi_1) + \delta, \psi_G(\delta)\}$, for R = N, B, X. Since $S_O^B(\varphi_1) < S_O^N(\varphi_1) = S_O^X(\varphi_1)$, it is clear that type (G, δ) is best off in regime B. Finally, consider the equilibrium payoff to type (I, δ) under the various regimes. Let $\psi_I(\delta) \equiv (1 - \eta_2)(1 - \eta_3)\pi(S_T + \delta) + c_2 + (1 - \eta_2)c_T$ denote this type's expected payoff from rejecting the plea offer in regime N. At the time of evaluation, I does not know whether P will become informed (which will happen with probability γ) or remain uniformed (with probability $1 - \gamma$). Type (I, δ)'s regime-dependent payoffs are:

Regime N: $\gamma\min\{(S_O^N(E_1) + \delta)(1 - \eta_3), \psi_I(\delta)\} + (1 - \gamma)\min\{(S_O^N(\varphi_1) + \delta)(1 - \eta_3), \psi_I(\delta)\}$.

Regime B: $\min\{(S_O^B(\varphi_1) + \delta)(1 - \eta_3), (1 - \gamma)\psi_I(\delta)\}$.

Regime X: $(1 - \gamma)\min\{(S_O^X(\varphi_1) + \delta)(1 - \eta_3), \psi_I(\delta)\}$.

Since $S_O^X(\varphi_1) = S_O^N(\varphi_1)$, it is clear that (I, δ) prefers regime X to regime N for all δ. We can re-write the regime X payoff as follows:

Regime X: $\min\{(1 - \gamma)(S_O^X(\varphi_1) + \delta)(1 - \eta_3), (1 - \gamma)\psi_I(\delta)\}$.

It is then clear that a sufficient condition for (I, δ) to prefer regime X to regime B for all δ is that $(1 - \gamma)S_O^X(\varphi_1) < S_O^B(\varphi_1)$. If this inequality holds, then $(1 - \gamma)(S_O^X(\varphi_1) + \delta) < (S_O^B(\varphi_1) + \delta)$ for all δ since the expression on the left-hand-side increases at the rate $1 - \gamma$ whereas the right-hand-side increases at the rate 1. Recall that (for fixed γ) $S_O^B(\varphi_1)$ converges to $S_O^X(\varphi_1)$ as λ_0 goes to zero. Thus, for λ_0 sufficiently small, a D of type (I, δ) most prefers regime X for all δ.