Vanderbilt University Department of Economics Working Papers 14-00006

Informal Sanctions on Prosecutors and Defendants and the Disposition of Criminal Cases

Andrew F. Daughety
Vanderbilt University

Jennifer F. Reinganum
Vanderbilt University

Abstract

We model the strategic interaction between a prosecutor and a defendant when informal sanctions by outside observers (society) may be imposed on both the defendant and the prosecutor. Outside observers rationally use the disposition of the case (plea bargain, case drop, acquittal, or conviction) to impose these sanctions, but also recognize that errors in the legal process (as well as hidden information) means they may misclassify defendants and thereby erroneously impose sanctions on both defendants and prosecutors. If third parties prefer a legal system with minimal regret arising from classification errors, there is a unique equilibrium wherein the guilty defendant accepts the prosecutor's proposed plea offer with positive probability, the innocent defendant rejects the proposed offer, and the prosecutor chooses to take all defendants who reject the offer to trial. We also consider the effect of increasing the informativeness of the jury's decision by extending the model to allow for a three-verdict outcome (not guilty, not proven, and guilty; the “Scottish” verdict). We find that: 1) guilty defendants are worse off, as plea bargains get tougher but the rejection rate does not change; 2) innocent defendants are better off; 3) the prosecutor's overall payoff goes up; and 4) the outside observers' regret over possible misapplication of informal sanctions is reduced. Thus the Scottish verdict is justice-improving when compared with the standard (two-outcome) verdict.

This paper was partly-written while visiting at the Paris Center for Law and Economics (Daughety) and the University of Paris 2 (Reinganum); we especially thank Bruno Deffains for providing a supportive research environment. We also thank Scott Baker, David Bjerk, Richard Boylan, Rosa Ferrer, Francoise Forges, Luigi Franzoni, and seminar participants at the University of Bologna for comments and suggestions on an earlier version.
Citation: Andrew F. Daughety and Jennifer F. Reinganum, (2014) "Informal Sanctions on Prosecutors and Defendants and the Disposition of Criminal Cases", *Vanderbilt University Department of Economics Working Papers*, VUECON-14-00006.
Contact: Andrew F. Daughety - andrew.f.daughety@vanderbilt.edu, Jennifer F. Reinganum - jennifer.f.reinganum@vanderbilt.edu.
Submitted: July 05, 2014. **Published:** July 08, 2014.
URL:http://www.accessecon.com/Pubs/VUECON/VUECON-14-00006.pdf

Informal Sanctions on Prosecutors and Defendants and the Disposition of Criminal Cases*

by

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

Original: May 2014
This version: June 2014

* This paper was partly-written while visiting at the Paris Center for Law and Economics (Daughety) and the University of Paris 2 (Reinganum); we especially thank Bruno Deffains for providing a supportive research environment. We also thank Scott Baker, David Bjerk, Richard Boylan, Rosa Ferrer, Francoise Forges, Luigi Franzoni, and seminar participants at the University of Bologna for comments and suggestions on an earlier version.

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

Informal Sanctions on Prosecutors and Defendants and the Disposition of Criminal Cases

Abstract

We model the strategic interaction between a prosecutor and a defendant when informal sanctions by outside observers (society) may be imposed on both the defendant and the prosecutor. Outside observers rationally use the disposition of the case (plea bargain, case drop, acquittal, or conviction) to impose these sanctions, but also recognize that errors in the legal process (as well as hidden information) means they may misclassify defendants and thereby erroneously impose sanctions on both defendants and prosecutors. If third parties prefer a legal system with minimal regret arising from classification errors, there is a unique equilibrium wherein the guilty defendant accepts the prosecutor's proposed plea offer with positive (but fractional) probability, the innocent defendant rejects the proposed offer, and the prosecutor chooses to take all defendants who reject the offer to trial.

We also consider the effect of increasing the informativeness of the jury's decision by extending the model to allow for a three-verdict outcome (not guilty, not proven, and guilty), sometimes referred to as the "Scottish" verdict. We find that: 1) guilty defendants are worse off, as plea bargains get tougher but the rejection rate does not change; 2) innocent defendants are better off; 3) the prosecutor's overall payoff goes up; and 4) the outside observers' regret over possible misapplication of informal sanctions is reduced. Thus the Scottish verdict is justice-improving when compared with the standard (two-outcome) verdict.

1. Introduction

All of us are familiar (if only from TV and the movies) with the fact that the criminal justice process provides formal sanctions for convicted defendants. These formal sanctions generally take two forms: incarceration (with the possibility of probation being used in some cases) and fines. In this paper we consider a third form of sanction that arises from members of society who observe pieces of the process, draw conclusions about the participants, and impose costs on the (perceived) offending party; we refer to these as informal sanctions.

Formal sanctions are imposed on defendants who are judged guilty (that is, who have accepted a plea bargain or were found guilty at trial),[1] and informal sanctions on convicted defendants have a long history; defendants who have been judged guilty and served their sentences may find it difficult to find housing and employment after release. Informal sanctions can also fall on defendants who have only been arrested, and for whom any charges have been dropped. Only one-fourth of the states actually prohibit the use of (pure) arrest information by employers when hiring (and the degree of enforcement is unclear).[2] Many states are silent on such matters (leaving the use of such information for hiring purposes entirely at the discretion of employers), while the remainder have imposed some limitations. For example, while Michigan prohibits employers asking about misdemeanor arrests that did not lead to conviction, no restrictions are placed on asking about felony arrests that did not lead to conviction. There are a number of firms that specialize in investigating job candidates' past criminal records (which typically include arrests, even if those

[1] We abstract from the use of formal sanctions for officials, but abuses such as prosecutorial misconduct can lead to formal sanctions. A fairly well-known example of prosecutorial misconduct involved Michael Nifong, the District Attorney for Durham County, NC, who was disbarred for his actions in the 2006 Duke University lacrosse case prosecution; see *NC State Bar v. Nifong* (June 16, 2007). Nifong also was convicted of criminal contempt of court for lying to a judge and served one day in jail and paid a $500 fine; see Associated Press (2007).

[2] The website Nolo.com provides state-level detail on the state and federal restrictions on the use by employers of information about arrest or conviction (www.nolo.com/legal-encyclopedia/state-laws-use-arrests-convictions-employment.html; accessed June 24, 2014). The Equal Opportunity Employment Commission provides guidance on what could constitute discriminatory hiring from a federal perspective, and only prohibits blanket policies of not hiring those with arrest records. The EEOC reports survey results that 92% of responding employers use criminal or background checks on all or some of job candidates (http://www.eeoc.gov/laws/guidance/arrest_conviction.cfm#IIIA; accessed June 24, 2014).

arrests did not lead to conviction) and provide such services to employers. A recent online development has been websites that publish booking photos ("mug shots") that are part of the public record.[3] Moreover, defendants who have been acquitted may be greeted with the suspicion that they were actually guilty but the jury was unable to formally reach that conclusion.[4] Informal sanctions may also be applied to officials in the system; for concreteness we specifically focus on prosecutors, but others may also be subject to such sanctions. Informal sanctions on officials may take the form of voters casting ballots against those who face reelection to office, or decisions by higher-level officials to not promote (or to fire) particular prosecutors.

Thus, both defendants and prosecutors may experience losses due to informal sanctions, no matter what the outcome for a specific case. We view formal sanctions as operating via the existing judicial system and limit ourselves to incarceration, while informal sanctions come from members of society and reflect the beliefs of such "outside observers." Importantly, we show how informal sanctions may affect (including limit) the use of formal sanctions.

More precisely, informal sanctions for the defendant involve outside observers drawing an inference about how likely it is that the defendant is guilty, given the case disposition, and applying sanctions that are proportional to this belief. These sanctions correspond to the outside observers withdrawing from further interactions with the defendant; for instance, they may choose not to hire

[3] See the discussion of the case of Dr. Janese Trimaldi, among others, in Segal (2013). Despite the fact that all charges against her were dropped, her booking photo (which is a public record) began to appear at online mug-shot sites. Segal estimates that there are over 80 such sites that generally charge people to remove the images; he indicates that fees for removal of information tend to run between $30 and $400 and, since multiple sites may post the picture, the cost of eliminating this information from the web can be exorbitant.

[4] One juror in the case against Casey Anthony (acquitted of murdering her two-year-old daughter), stated: "I did not say she was innocent; I just said there was not enough evidence. If you cannot prove what the crime was, you cannot determine what the punishment should be." See Burke, et. al. (2011).

him for a job, or to avoid social interactions with him, and so on.[5] The proportional specification is a simple way to ensure that the defendant suffers worse informal sanctions the higher is the outside observers' belief in his guilt. Although these informal sanctions are costly to the defendant, we assume that the outside observers do not bear any direct cost of imposing them; for instance, they can simply hire, or interact socially with, someone else. We also assume that the outside observers impose informal sanctions on the prosecutor, in proportion to their posterior belief that she has punished an innocent defendant (through conviction at trial or through a plea-bargained conviction), or failed to punish a guilty defendant (either through acquittal at trial or through dropping the case). Informal sanctions on the prosecutor can affect her career concerns via election, appointment, promotion, or selection for judgeships, or outside opportunities in private law firms and universities.

These informal sanctions will affect both the feasibility and the optimality of plea bargaining. In particular, we find that the only type of equilibrium that can exist is a semi-separating one wherein innocent defendants reject the plea offer, whereas guilty defendants mix between accepting and rejecting the plea offer. Because the prosecutor has the option to drop the case, there must be a sufficient fraction of guilty defendants among those that reject the plea offer in order to incentivize the prosecutor to go to trial following rejection. Informal sanctions may restrict the feasibility and optimality of the equilibrium wherein at least some defendants settle, as (in equilibrium) accepting the plea bargain results in a clear inference of guilt, which results in the highest informal sanction against the defendant. If the informal sanction rate for the defendant is too high, then it will not be

[5] This is similar to Iacobucci's (2014, p. 190) reputational sanction that "... arises because observers have changed their views about the benefits of dealing with a wrongdoer that has revealed by its wrong its type as one that is unattractive to trading partners (with 'trading' conceived broadly)." He assumes that there are two types of agents (good and bad) and two actions (comply and not comply) and characterizes when separating and pooling equilibria exist, noting (p. 196) that there can be multiple equilibria due to "an interaction between observers' expectations and firm behavior." Many other papers consider how formal and informal sanctions interact in generating deterrence; see Iacobucci (2014) and the references therein. We do not address the issue of deterrence in this paper, but rather focus on plea bargaining.

possible to induce a defendant to accept a plea bargain. Similarly, a prosecutor may prefer to take a case to trial rather than settling via a plea bargain if the informal sanction rate on the defendant is too high, because the prosecutor has to discount the formal sanction (in the plea offer) in order to induce the defendant to accept both the plea offer and the informal sanction that results when he thereby reveals his guilt. The threshold informal sanction rate on the defendant increases in the informal sanction rate on the prosecutor for punishing the innocent, as a higher informal sanction rate on the prosecutor for punishing an innocent person makes trial (which can result in erroneous convictions) less attractive to the prosecutor. Thus, some informal sanctions work in opposition to others.

Because the equilibrium fraction of guilty defendants among those that reject the plea offer is co-determined with the fraction that outside observers expect to reject the plea offer, there is a continuum of semi-separating equilibria, indexed by this fraction. There is a smallest equilibrium fraction that is necessary to incentivize the prosecutor to go to trial following a rejection (rather than dropping the case). We show that, if outside observers prefer to minimize the extent of erroneously-imposed informal sanctions, then they prefer the equilibrium wherein the smallest fraction of guilty defendants reject the plea offer. That is, they prefer the equilibrium which entails the greatest amount of successful plea bargaining. This equilibrium involves the lowest possible amount of "misclassification" because those that accept the plea offer are revealed to be guilty types, and trial is the clearest possible signal of innocence (subject to the noise that is required to incentivize the prosecutor to go to trial following a rejected plea offer).

Finally, we consider a legal system wherein the outside observers are able to acquire more information from a jury as to the degree of guilt of the defendant. Specifically, we extend the model to consider the "Scottish verdict" wherein the verdict allows for three outcomes: not guilty, not proven, and guilty. The intermediate case, not proven, carries no formal sanctions (it is a form of acquittal), but represents an outcome wherein jurors felt that the prosecution's case against the defendant was insufficiently strong to meet the high evidentiary standard needed in a criminal case (beyond a reasonable doubt), but also reflects an unwillingness on the part of the jury to assert that the defendant was not guilty. We show that this finer resolution of the jury's assessment leads to increased expected costs to a truly guilty defendant, lower expected costs to a truly innocent defendant, and informal sanctions by outside observers that are more likely to be deserved; altogether, these results suggest that the Scottish verdict is likely to be justice-enhancing relative to the standard two-outcome (guilty/not guilty) verdict.

Related Literature

Landes (1971) provides a complete-information model wherein the prosecutor's objective is to maximize expected sentences obtained from a collection of defendants, subject to a resource constraint; the potential for innocent defendants is not considered. Grossman and Katz (1983) and Reinganum (1988) provide screening and signaling models, respectively, of plea bargaining wherein the prosecutor maximizes a utility function that corresponds to social welfare. Grossman and Katz assume that the defendant knows whether he is guilty or innocent; the prosecutor's plea offer screens the defendant types so that the innocent go to trial whereas the guilty accept the plea offer. Reinganum assumes that the defendant knows whether he is guilty or innocent, but the prosecutor also has private information: she knows the actual likelihood of conviction at trial. In this case, the plea offer signals the strength of the prosecutor's case but it does not screen the defendant types; both guilty and innocent defendants randomize between accepting the plea offer and going to trial.

Thus, the prosecutor goes to trial against a mixture of guilty and innocent defendants. In both of these models, however, it is assumed that the prosecutor is committed to taking the case to trial following a rejected plea offer; this means that even if the prosecutor knew the defendant was innocent, she would pursue a conviction.

Absent this commitment to trial following a rejected plea, a putative equilibrium in which the guilty accept the plea offer and the innocent reject it is undermined by the prosecutor's desire to drop the case rather than proceeding to trial against a defendant that (she now believes) is innocent. Franzoni (1999) and Baker and Mezzetti (2001) explicitly incorporate a credibility constraint[6] into a screening model, which requires that a sufficiently high fraction of guilty defendants reject the plea offer.[7] We also incorporate this kind of credibility constraint; our model is closest in terms of the prosecutor's payoff functions to that of Baker and Mezzetti because in both models the prosecutor faces a risk of convicting an innocent defendant.[8] However, we will incorporate informal sanctions that fall on both the defendant and the prosecutor, depending on the disposition of the case (e.g., convicted; acquitted; plea-bargained; or dropped).[9]

The papers discussed thus far (as well as our paper) assume that the judge or jury makes its

---

[6] See Nalebuff (1987) for a screening model with a possibly-binding credibility constraint in the case of a civil suit.

[7] In Franzoni's model, innocent defendants are never convicted, so the prosecutor simply maximizes the expected penalty imposed on the guilty less the cost of the effort she expends to investigate prior to trial. However, if only innocent defendants reject the plea offer, then the prosecutor is unwilling to expend any effort on investigation; thus, equilibrium must involve some guilty defendants rejecting the plea offer as well.

[8] In Baker and Mezzetti's model, a prosecutor obtains a payoff of x (resp., $-x$) if a guilty (resp. innocent) defendant gets a sentence of x. The prosecutor obtains a payoff of zero if she frees an innocent defendant and $-\alpha x$ if she frees a guilty defendant. Finally, the prosecutor does not have a cost of trial, but loses the amount c whenever she loses at trial (this is viewed as a reputational cost). Thus, in their model the prosecutor has internal concern for punishing the innocent and letting the guilty go free, and they obtain a unique semi-separating equilibrium. In our model these sanctions are provided by the outside observers, and we obtain a continuum of semi-separating equilibria, and then use a selection criterion to obtain a unique (selected) equilibrium.

[9] Baker and Mezzetti also include a stage following plea bargaining where a public signal can provide exonerating evidence and the prosecutor can drop the case against an exonerated defendant, followed by a private signal observed only by the judge or jury that determines whether the defendant is convicted or acquitted. We omit this public signal stage (which is not necessary to the equilibrium in Baker and Mezzetti, but allows the prosecutor to avoid trying some innocent defendants), but we retain the private signal that is observed only by the judge or jury, which is crucial to the screening aspect of the model.

decision based only on the evidence ("signal") they observe in the course of the trial and the specified standard of proof. That is, they follow their instructions rather than acting as rational Bayesian agents. Bjerk (2007) provides a model in which the jury acts as a rational Bayesian agent, and this undermines an equilibrium wherein the prosecutor screens defendant types perfectly. For if the prosecutor was expected to settle with all of the guilty defendants, then the jury would rationally infer that those coming to trial must be innocent, and the jury would acquit (but then the guilty would refuse to settle). Rather than the prosecutor herself wanting to drop the case instead of proceeding against a defendant that (she believes) is innocent (the prosecutor is not allowed to drop the case in Bjerk's model), it is the jury that wants to acquit a defendant that (it believes) is innocent. The beliefs of the jury are self-fulfilling and Bjerk finds that the model can have a continuum of equilibria, indexed by the evidentiary threshold needed for the jury to convict the defendant.[10]

Our model also has a continuum of equilibria, but these are based on the (rational Bayesian) beliefs of the outside observers, who impose informal sanctions on both the defendant and the prosecutor. As in Franzoni (1999), Baker and Mezzetti (2001), and Bjerk (2007), our prosecutor will face a credibility constraint in that she will have to ensure that there are enough guilty defendants among those that reject the plea offer to rationalize her going to trial. Our multiple equilibria are indexed by the fraction of guilty defendants among those that reject the plea offer (again, innocent defendants always reject the plea offer). When outside observers believe that this fraction is high, then they impose high informal sanctions on defendants following a trial, which

[10] Bjerk assumes that both the prosecutor and the jury maximize social welfare (given their respective beliefs and information), and trials are costless. Our prosecutor's objective increases in the sentences she obtains, but is also influenced by informal sanctions imposed by outside observers based on her perceived errors. Bjerk does not argue for a particular selection from among the equilibria he identifies. We argue for a particular equilibrium to be selected based on the desire of outside observers to minimize the extent of informal sanctions that they impose in error.

allows the prosecutor to make a high plea offer that is rejected by a high fraction of guilty defendants. On the other hand, if the outside observers believe that the fraction of guilty defendants among those that reject the plea offer is low, then they impose low informal sanctions on defendants following a trial, which constrains the prosecutor to make only a low plea offer that is rejected by a low fraction of guilty defendants. Because the outside observers' beliefs are based on coarse information (i.e., the case disposition), they will sometimes impose informal sanctions on defendants and prosecutors that are excessive or insufficient. We select among the equilibria in our model by positing a preference on the part of the outside observers to minimize the expected regret due to erroneously-imposed informal sanctions.[11]

Finally, as indicated above, there are a number of different formulations for the prosecutor's objective, ranging from expected sentences to social welfare to a mixture of motivations. Prosecutors are in an unusual position as they are supposed to represent society but they will clearly have personal preferences and career concerns as well. The general issue of what it is that prosecutors are maximizing is important to formulating models of plea bargaining. Boylan and Long (2005) used data on young federal prosecutors and found that those assistant U.S. attorneys in districts with very high private salaries were more likely to take cases to trial, and viewed this as evidence that those positions for young prosecutors were sought by individuals who wanted the trial experience, with an eye towards an eventual private-sector job. Boylan (2005) expands on this by examining the careers of U.S. attorneys and finds that the length of prison sentences obtained (but not conviction rates) is positively related to positive outcomes in their career paths.

[11] Defendants also prefer this equilibrium. As we discuss later, despite the fact that outsiders incur no cost for imposing sanctions, an alternative basis for choosing this equilibrium is that outsiders (recognizing that they may be defendants one day) under a veil of ignorance would prefer the same equilibrium as that which minimizes expected regret.

Plan of the Paper

In Section 2, we provide the notation and formal model. In Section 3, we describe the equilibria of the model (the equilibrium concept will be Perfect Bayesian equilibrium), and a rationale for selecting among them. We then discuss some important comparative statics due to the informal sanctions on the defendant and on the prosecutor. Section 4 extends the model to the Scottish verdict and shows that this refinement enhances justice. In Section 5, we provide a summary and some additional discussion, and we raise some possible extensions. A few of the most salient technical issues are included in the Appendix while a Technical Appendix[12] contains the full details of the analysis.

2. Modeling Preliminaries

Description of the Game

Our game commences after the arrest of the defendant. The defendant, D, will be taken to be male, and the prosecutor, P, female. The exogenous parameters of the game include the sentence upon conviction (S_c), the evidentiary criterion used by the jury for conviction (γ_c), and the cost of trial for each agent (k^P for P and k^D for D). More detail on the notation (and the informal sanctions, which also have exogenously-determined elements) will be provided as we progress, but a basic notational convention will be that outcomes or actions appear as subscripts while "ownership" – that is, which agent is affected by the variable or parameter of interest – is indicated by a superscript.

There are five stages in the game:

Stage 1: Nature (N) draws D's type, denoted by t, and this is revealed to D only.

[12] Available at http://www.vanderbilt.edu/econ/faculty/Daughety/DR-InformalSanctionsandCaseDispositions-TechApp.pdf

Stage 2: P makes a plea bargain offer of $S_b \geq 0$ (since P is the only agent who can make such offers in our analysis, ownership is implied).

Stage 3: D chooses whether to accept (A) or reject (R) the plea bargain offer; if he accepts the offer (outcome b), the game ends and payoffs (π_b^P and π_b^D) are obtained.

Stage 4: If D has chosen R, then P now chooses whether to drop the case (outcome d) or pursue it to trial (action T). A case that is dropped means that P and D obtain payoffs π_d^P and π_d^D, respectively.

Stage 5: If the case goes to trial (T), then Nature (N) draws the evidence of guilt, e, and the jury (J) uses the rule that if $e \geq \gamma_c$, then the outcome is conviction (c), while otherwise the outcome is acquittal (a). In the case of conviction P and D obtain π_c^P and π_c^D, respectively, while in the case of acquittal P and D obtain π_a^P and π_a^D, respectively.

Figure 1 below illustrates the sequence of actions and outcomes in the game (the "fans" indicate that the specified variable can take on a continuum of values); we have not indicated information sets so as to make the diagram more readable.

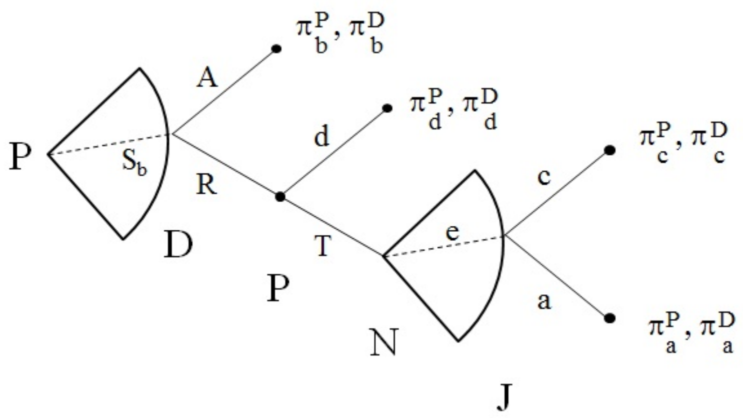

Figure 1: Sequence of Actions and Case Dispositions

Let $t \in \{I, G\}$ denote D's privately-known type (Innocent or Guilty), and let $\lambda > 0$ denote the fraction of innocent Ds; that is $\lambda = \Pr\{t = I\}$, the probability that a D at the left end of Figure 1 is innocent of the crime for which he was arrested. This parameter reflects some initial level of evidence gathered by the police. As indicated above, no further evidence is available until P and D go to trial, in which case a draw of evidence of guilt, $e \in [0, 1]$, occurs; this draw is influenced by the underlying type for D and is observed only by the jury. The jury is instructed to convict if its evidence signal exceeds a threshold (γ_c) and to otherwise acquit the defendant. We denote the distribution of evidence (given D's type) as $F(e \mid t)$. Since for any evidentiary standard for conviction, γ_c, the jury (J) will choose outcome a when $e \leq \gamma_c$, then for either type t, $F(\gamma_c \mid t)$ is the probability that D is acquitted and $1 - F(\gamma_c \mid t)$ is the probability that D is convicted. This motivates the assumption that at any level of evidence e, the probability of acquittal for an innocent D is higher than that for a guilty D:

$$F(e \mid I) > F(e \mid G) \text{ for all } e.$$

Finally, to aid readability in writing out payoffs, let F_t denote $F(\gamma_c \mid t)$, for $t = I, G$, so our assumption becomes $F_I > F_G$.

The sentences S_c and S_b are formal sanctions. Informal sanctions are penalties imposed by outside observers on both defendants and prosecutors; to reduce the verbiage, let Θ denote the outside observer(s).[13] Informal sanctions are based on Θ's beliefs, which are contingent on the case disposition (a, b, c, or d). We assume that these informal sanctions are proportional to the observers' beliefs, which depend upon the inferred type of defendant and the observed outcome of the legal

[13] Since we will not be accounting for any heterogeneity among the outside observers, we will refer to both singular possessive and plural possessive Θ (e.g., we will refer to Θ's and Θs' beliefs) interchangeably. Outside observer beliefs will always be denoted by μ, so ownership of such beliefs will be obvious and therefore superscripting them with a Θ is unnecessary.

process.[14] More precisely, these beliefs represent Θ's posterior probability that the defendant is type t, given the case disposition was y, and are denoted $\mu(t \mid y)$, where t = I, G and y = a, b, c, d. Note that this also means that Θ cannot directly observe the plea offer, S_b, the levels of P's and D's payoffs, or the evidence draw e.[15]

Informal sanctions imposed on D are of the form $r^D\mu(G \mid y)$, where $r^D \geq 0$ is an exogenous parameter.[16] That is, given any case disposition, Θ assesses the posterior likelihood that D is guilty, and then imposes informal sanctions at the rate r^D. These informal sanctions, which are increasing in the posterior assessment of guilt, reflect the fact that Θs may be future trading partners (broadly-construed) who decline to trade with the defendant; we assume that the observers themselves do not suffer a direct cost of imposing these sanctions (though we will later assume that they prefer a society that minimizes the extent of erroneously-imposed informal sanctions).

As indicated earlier, Θs also impose informal sanctions on P, reflecting the notion that errors occur within the legal process; informal sanctions imposed on the prosecutor arise when there is a belief that prosecutors should be blamed for such perceived errors; for instance, a guilty defendant may be acquitted or the case may be dropped. In these instances, P has allowed a guilty D to escape punishment. The associated informal sanctions are given by $r_G^P\mu(G \mid y)$, for $y \in \{a, d\}$. On the other hand, an innocent defendant can be convicted or may accept a plea bargain. In these instances, the prosecutor has punished an innocent defendant. The associated informal sanctions are given by

[14] For simplicity, we assume that the outside observers always observe the case disposition. However, it is trivial to allow this to occur only with positive probability. Probabilistic observation would simply re-scale the informal sanction rates by pre-multiplying these rates by the probability that the observers actually do observe the case disposition.

[15] It is very plausible that Θ would not observe a rejected plea offer. Our analysis assumes that Θ does not observe the plea bargain offer S_b, even if D accepts the bargain and the fact of that acceptance is observed. We speculate that, due to the structure of the game and the fact that there are only two types of D, observing S_b if D accepts the offer would not affect Θ's out-of-equilibrium beliefs, but we leave this as an item for future research; see our discussion of transparency in the last section.

[16] In general we think of this rate as positive, but it could be negative, such as might hold if D was a gang member desiring "street cred."

$r^P_I\mu(I \mid y)$, for $y \in \{b, c\}$. We assume that r^P_I and r^P_G are non-negative. While not needed in the model, a natural assumption would be that the sanction rate for the prosecutor is higher when an innocent defendant is punished than when a guilty defendant is not punished; that is, $r^P_I > r^P_G$.

D's Payoffs

We are interested in non-cooperative solutions for the game that exhibit sequential rationality by G, I, P and Θ, so we will first develop payoff functions starting from the outcomes (a, b, c, and d). Since trial ends in conviction or acquittal, D's payoffs on the right-hand-side of Figure 1 can be written as (note that D chooses actions so as to minimize his total expected payoff):

$$\pi^D_c = S_c + k^D + r^D\mu(G \mid c); \tag{1a}$$

and

$$\pi^D_a = k^D + r^D\mu(G \mid a). \tag{1b}$$

That is, going to trial costs D the amount k^D. Conviction results in the formal sanction S_c plus the informal sanction $r^D\mu(G \mid c)$; since I-types may have been convicted (the evidence draw for an I-type could, conceivably, result in conviction), Θ recognizes that conviction is not a guarantee of guilt, so $\mu(G \mid c)$ will be less than one. Similarly, acquittal generally does not imply innocence, so Θ's belief $\mu(G \mid a)$ will be positive and D will bear both the cost of court and the informal sanction $r^D\mu(G \mid a)$.

We can write D's payoff (given his type t) from going to trial as the weighted combination of the elements in equations (1a) and (1b), where the weights reflect the likely outcome at trial:

$$\pi^D_T(t) = S_c(1 - F_t) + k^D + r^D\mu(G \mid c)(1 - F_t) + r^D\mu(G \mid a)F_t,\ t \in \{I, G\}. \tag{2}$$

For instance, if t = I, then if D goes to trial, he expects to be convicted (outcome c) with probability $1 - F_I$, in which case he will receive the formal sanction S_c and the informal sanction $r^D\mu(G \mid c)$. He expects to be acquitted (outcome a) with probability F_I, in which case he will receive no formal sanction but Θ still believes there is a chance D is guilty despite his acquittal, and imposes the informal sanction $r^D\mu(G \mid a)$. Also note that D pays his trial costs k^D regardless of the trial outcome. The following result is shown in the Technical Appendix.

Remark 1: $\pi_T^D(I) < \pi_T^D(G)$.

That is, a D of type I has a lower expected loss from trial than does a D of type G. This result follows from: 1) the linear structure of $\pi_T^D(t)$ and 2) the fact that $F_I > F_G$. Remark 1 is important in that it suggests that the equilibrium might involve full screening (wherein, say, P makes an offer that only G-types accept and I-types reject), or partial screening (wherein, say, P makes an offer that all of one type reject, and that some of the other type reject); that is, properly-constructed offers in the plea bargaining stage may yield information about D's type. We return to this in Section 3 in our discussion of the equilibria of the game.

If P offers a plea bargain of S_b, then D can choose to accept (A) or reject (R) the offer. D's payoff from accepting a plea bargain of S_b is:

$$\pi_b^D = S_b + r^D\mu(G \mid b); \tag{3}$$

that is, he receives the formal sanction S_b plus the informal sanction that observers impose because, having accepted the plea offer (outcome b), they believe that he is guilty with probability $\mu(G \mid b)$.

Similarly, D's payoff if P drops the case is:

$$\pi_d^D = r^D\mu(G \mid d), \tag{4}$$

which reflects Θ's beliefs that D might have been guilty.

Since P may mix between going to trial and dropping the case following a rejection of the plea offer by D, let ρ^P denote the probability that P takes the case to trial following rejection by D. Combining equations (2) and (4), weighted by the probability that P takes the case to trial, yields

D's expected payoff following rejection (given his type) as:

$$\pi^D_R(t) = \rho^P \pi^D_T(t) + (1 - \rho^P)\pi^D_d. \tag{5}$$

P's Payoffs

Again, starting at the right of Figure 1, since trial ends in conviction or acquittal, P's payoffs on the right-hand-side of Figure 1 can be written as (note that P chooses actions so as to maximize her total expected payoff):

$$\pi^P_c = S_c - k^P - r^P_I \mu(I \mid c); \tag{6a}$$

$$\pi^P_a = -(k^P + r^P_G \mu(G \mid a)). \tag{6b}$$

Next we obtain P's expected payoff from going to trial; this turns out to be somewhat more complicated than D's corresponding payoff because P and Θ have different amounts of information on which to form beliefs. When the prosecutor makes the plea offer S_b, she does not know whether the defendant is guilty or innocent, so D's decision to accept or reject the offer will affect the prosecutor's posterior belief that he is guilty. The prosecutor's beliefs may differ from those of the observer because she observes the plea offer, whereas the observer observes only the disposition of the case. To capture this, let $\nu(G \mid R)$ (resp., $\nu(G \mid A)$) denote the prosecutor's posterior probability that the defendant is guilty,[17] given that he rejected (resp., accepted) the plea offer S_b. Of course, in equilibrium, P's beliefs and Θ's beliefs must be the same (and must be correct).

The prosecutor's payoff from going to trial (given her beliefs following the defendant's rejection of her plea offer) can be written as:

$$\pi^P_T = \nu(G \mid R)\{S_c(1 - F_G) - k^P - r^P_I \mu(I \mid c)(1 - F_G) - r^P_G \mu(G \mid a)F_G\}$$
$$+ \nu(I \mid R)\{S_c(1 - F_I) - k^P - r^P_I \mu(I \mid c)(1 - F_I) - r^P_G \mu(G \mid a)F_I\}. \tag{7}$$

[17] As with Θ's ownership of the beliefs μ, P's ownership of her beliefs ν will not be superscripted, so as to avoid unnecessary clutter. P's beliefs will also depend on the plea offer S_b, but this would needlessly complicate the notation so this dependence is suppressed.

This is interpreted as follows. Given rejection of the plea offer, P believes that D is guilty with probability $\nu(G \mid R)$, in which case she expects a conviction with probability $1 - F_G$ and an acquittal with probability F_G. If D is convicted, P obtains utility from the formal sanction S_c but observers still harbor the posterior belief $\mu(I \mid c)$ that D may be innocent (despite his conviction), and impose on P the informal sanction $r_I^P\mu(I \mid c)$. If D is acquitted, then the Θs still harbor the posterior belief $\mu(G \mid a)$ that D is guilty (despite his acquittal) and impose on P the informal sanction $r_G^P\mu(G \mid a)$. Regardless of the case disposition (a or c), P pays the trial costs k^P. The second part of P's payoff, wherein she believes that D is innocent with probability $\nu(I \mid R)$, is interpreted similarly.

If P's plea offer is accepted then she obtains the following payoff:

$$\pi_b^P = S_b - r_I^P\mu(I \mid b). \tag{8}$$

Equation (8) indicates that P's payoff if the offer is accepted is the level of the plea offer minus an informal sanction imposed on P by Θ that reflects Θ's belief in the possibility that an innocent D accepted the offer.

Following rejection of the plea offer, P has the option to drop the case. If she does so, then she receives no payoff from formal sanctions, but she receives an informal sanction from Θ, who believes with probability $\mu(G \mid d)$ that D is guilty, so by dropping the case P let a guilty defendant go free. Thus, P's payoff from dropping the case is simply:

$$\pi_d^P = -\, r_G^P\mu(G \mid d). \tag{9}$$

As earlier, since P may mix between dropping the case and going to trial, P's expected payoff following a rejection by D is given by:

$$\pi_R^P = \rho^P\pi_T^P + (1 - \rho^P)\pi_d^P. \tag{10}$$

3. Results

In this section we provide the main results; a sketch of the derivation is in the Appendix while the Technical Appendix contains the complete analysis. We start by providing notation for the strategies for each type of D and for P, for Θ's conjectures about these strategies, and two restrictions on the parameter space that reflect sensible behavior by P. We then describe the game's equilibria and the reasons for selecting a specific equilibrium.

First, we denote a strategy profile as a four-tuple of strategies, one for P in Stage 2, one for each type of D in Stage 3, and one for P in Stage 4:

$$(S_b, \rho_G^D, \rho_I^D, \rho^P) \in [0, \infty)\times[0, 1]\times[0, 1]\times[0, 1],$$

where S_b is the plea offer made by P; ρ_G^D is the probability a D of type G rejects the plea bargain offer; ρ_I^D is the probability a D of type I rejects the plea bargain offer; and ρ^P is the probability that P goes to trial against a D who rejected the offer (rather than dropping the case). Thus, suppressing the plea offer for now, a candidate equilibrium wherein both types of D always reject the offer S_b and P always goes to trial is (1, 1, 1). Notice that P has conjectures about what the types of D will choose and D has conjectures about what P will do, conditional on D's action, and while Θ cannot observe P's and D's actions, Θ has conjectures about what P will do and about what the types of D will do. All conjectures must be correct in equilibrium, but since P's and D's conjectures are not the primary focus of the paper (and are pretty standard) we do not generate additional notation for these conjectures.

Let $(\rho_G^{D\Theta}, \rho_I^{D\Theta}, \rho^{P\Theta})$ denote Θ's conjectures about the strategies that will be used by G, I, and

P, respectively. For any such triple of conjectured strategies,[18] Θ's beliefs about D are as follows (we only provide the beliefs that D is of type G given an outcome; the corresponding beliefs for a D of type I are readily derivable):[19]

$$\mu(G \mid a) = \rho_G^{D\Theta}(1 - \lambda)F_G/[\rho_G^{D\Theta}(1 - \lambda)F_G + \rho_I^{D\Theta}\lambda F_I]; \tag{11a}$$

$$\mu(G \mid b) = (1 - \rho_G^{D\Theta})(1 - \lambda)/[(1 - \rho_G^{D\Theta})(1 - \lambda) + (1 - \rho_I^{D\Theta})\lambda]; \tag{11b}$$

$$\mu(G \mid c) = \rho_G^{D\Theta}(1 - \lambda)(1 - F_G)/[\rho_G^{D\Theta}(1 - \lambda)(1 - F_G) + \rho_I^{D\Theta}\lambda(1 - F_I)]; \tag{11c}$$

and

$$\mu(G \mid d) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \rho_I^{D\Theta}\lambda]. \tag{11d}$$

We employ Perfect Bayesian equilibrium and require that: 1) P maximizes her expected payoff by choosing her plea offer S_b, given Θ's conjectures, P's prior beliefs about D's type, and anticipating how the continuation game will play out following P's choice of plea offer; 2) each type of D minimizes his expected payoff by choosing his response to the plea offer, given Θ's conjectures, and anticipating how the continuation game will play out following his decision to accept or reject the plea offer; 3) P maximizes her expected continuation payoff via her choice to pursue trial or drop the case, given Θ's conjectures, and given P's posterior beliefs about D's type (based on his decision regarding the plea offer); and 4) all conjectures and beliefs are correct in equilibrium.

There are two scenarios concerning the decision by P whether to drop the case or go to trial that restrict the parameter space. First, if P (and Θ) know (or commonly believe) that D is innocent, then P should prefer dropping the case to going to trial. Second, imagine that both types of D were

[18] Technically Θ has a conjecture about S_b as well, but it is unneeded for the beliefs and we suppress this to avoid further clutter. Formally, the mathematical descriptions of Θ's beliefs presume that the strategy profile is fully-mixed, so that all nodes in the game are visited with positive probability, allowing us to use Bayes' Rule to provide the indicated formula. As we will see, (1, 1, 1) is an equilibrium of the game, so that in this equilibrium, the outcome b is an out-of-equilibrium outcome, and the value for $\mu(G \mid b)$ will need to be otherwise specified, since b will not be visited in equilibrium.

[19] P's strategy, ρ^P, does not affect the beliefs because it (or $1 - \rho^P$) multiplies each relevant numerator and denominator and thereby drops out of the analysis.

expected to reject P's offer. Then P's (and Θ's) posterior beliefs about the types would be the same as the prior beliefs (that $\Pr\{t = I\} = \lambda$); this would also be true if there was no stage involving P making a plea bargain offer. In this scenario, P should prefer trial over dropping the case; this will be true if the police arrest process procedure is sufficiently effective at discriminating between guilty and innocent persons; that is, if λ, the prior probability that D is of type I, is not "too large."[20] These two scenarios imply the following maintained restrictions.

Maintained Restriction 1 (hereafter, MR1): If P and Θ know (or commonly believe) that D is of type I, P should prefer to drop the case. Formally, this means that $\pi_T^P < \pi_d^P$ under the specified beliefs, which reduces to: $(S_c - r_I^P)(1 - F_I) - k^P < 0$. Notice that this can hold even if k^P is zero, as long as r_I^P exceeds S_c; lower values of r_I^P are still feasible if k^P is sufficiently positive.

Maintained Restriction 2 (hereafter, MR2): If P and Θ know (or commonly believe) that the fraction of type G among those that reject the plea offer is the same as the prior, then P should prefer to take the case to trial rather than dropping it. Formally, this restriction reduces to: $(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P] + \lambda[(S_c - r_I^P)(1 - F_I) - k^P] > 0$.

Note that the second term in brackets in MR2 is negative by MR1, so the above condition is an upper bound on λ (the arrest process is sufficiently effective). Moreover, MR1 implies that the first term in brackets in MR2, which represents the difference in the value of going to trial versus dropping the case against a D that is known (or commonly believed) to be type G, is positive. Basically, P prefers to drop rather than try cases against known innocent types, and prefers to try rather than drop cases against known guilty types; moreover, she still prefers to try rather than drop

[20] Essentially, this is the reason for requiring probable cause for an arrest (i.e., there is a reasonable basis to believe a potential D committed a specific crime).

against the prior mixture of defendant types.

Using these natural parameter restrictions, we find that the only[21] equilibria of the game involve the D of type I always rejecting the equilibrium plea offer (so, hereafter, $\rho^D_I = 1$), the D of type G rejecting the equilibrium plea offer with a positive probability ρ^D_G, and P never dropping a case when D rejects an offer. We provide a sketch of the proof of the following Proposition, which formalizes the description of the game's continuum of equilibria, in the Appendix (the complete proof is in the Technical Appendix). To simply the exposition of the Proposition, 1) let $\pi^D_T(G; \rho^D_G)$ be $\pi^D_T(G)$, as specified in equation (4), with Θ's beliefs evaluated at ρ^D_G and $\rho^D_I = 1$; and 2) define ρ^{D0}_G as the (unique) solution to the condition that $\pi^P_T = \pi^P_d$, where both of these expressions have Θ's beliefs and P's beliefs evaluated at ρ^{D0}_G (the specific value of ρ^{D0}_G appears in equation (14) below).[22]

Proposition 1: If r^D is not "too large" then there is a unique family of Perfect Bayesian equilibria for the game summarized by the four-tuple $(S_b(\rho^D_G), \rho^D_G, 1, 1)$, with $\rho^D_G \in [\rho^{D0}_G, 1]$, ρ^{D0}_G a positive fraction. For each ρ^D_G, P's equilibrium plea offer is $S_b(\rho^D_G) = \pi^D_T(G; \rho^D_G) - r^D$.

Note the following:

1) Both the D of type I and P use pure strategies in equilibrium: all Ds who are innocent reject P's equilibrium offer, and the equilibrium involves a sufficient fraction of Ds who are guilty to also reject the plea offer so that P will not choose to drop any case.[23]

2) One member of this family of equilibria is $(S_b(1), 1, 1, 1)$ wherein $\rho^D_G = 1$, in which case

[21] Alternative candidates for equilibria, such as fully-separating or fully-pooling candidates, or candidates involving type I accepting a plea offer, cannot be equilibria; see the Technical Appendix for details.

[22] Beliefs for Θ are given by equations 11(a-d), evaluated at arbitrary ρ^D_G and $\rho^D_I = 1$; beliefs for P are given by $\nu(G \mid R) = \rho^D_G(1 - \lambda)/[\rho^D_G(1 - \lambda) + \lambda]$, as ρ^D_G of the G-types and all of the I-types are expected to reject the plea offer.

[23] Out-of-equilibrium beliefs for Θ following an unexpected dropped case are $\mu(G \mid d) = \rho^D_G(1 - \lambda)/[\rho^D_G(1 - \lambda) + \lambda]$; since ρ^D_G of the G-types and all of the I-types are expected to reject the plea offer, Θ interprets the unexpected dropped case as an error on the part of P.

all Ds are rejecting P's offer, meaning that P's (and Θ's) beliefs about the types that chose R is the prior, so (by MR2) P goes to trial against all Ds.[24]

3) The smallest ρ_G^D-value (ρ_G^{D0}) in the family of equilibria is in (0, 1), so in almost all of the equilibria in the Proposition, G-types accept the plea offer with positive probability. From the perspective of Θ, since I-types always reject the offer, this means that $\mu(G \mid b)$ is one: a D who accepts the offer incurs the full sanction r^D from Θ.

Limits on Informal Sanctions

What do we mean by the qualifier in Proposition 1 "If r^D is not 'too large'"? Consider those equilibria wherein $\rho_G^D \in [\rho_G^{D0}, 1)$; that is, G-types accept P's offer with positive probability. In order for this to occur, P must: 1) be choosing S_b from a non-empty set and 2) not wish to defect by making a very high offer to D so as to make D reject the offer for sure. Notice that the Proposition indicates that $S_b(\rho_G^D) = \pi_T^D(G; \rho_G^D) - r^D$, so that the set of feasible S_b-values capable of inducing some acceptance is $[0, \pi_T^D(G; \rho_G^D) - r^D]$. Hence, in order to have a non-empty feasible set for S_b, it must be that $r^D \leq \pi_T^D(G; \rho_G^D)$. Substituting into this inequality yields our first condition restricting the level of informal sanctions, Condition 1 (see the Technical Appendix for more detail):

Condition 1 (feasibility). In order for P to be able to induce a D of type G to accept a plea offer, it must be that: $r^D \leq [S_c(1 - F_G) + k^D]/[1 - \mu(G \mid c)(1 - F_G) - \mu(G \mid a)F_G]$.

Multiplying through both sides of the above inequality by the expression in the denominator on the right, the resulting expression $r^D[1 - \mu(G \mid c;)(1 - F_G) - \mu(G \mid a;)F_G]$ is the increment in informal sanctions that the D of type G suffers by accepting a plea (which only a true G is expected

[24] Now we also need an out-of-equilibrium belief for Θ's observation of outcome b; that belief would be that the type is G, since G does worse at trial than does I, so observing b should be associated with t = G: $\mu(G \mid b) = 1$. Alternatively put, I is willing to reject the plea offer for a strictly larger probability of subsequently going to trial than is G, so selection of the "safe" option (accept) is attributed to G. This argument is associated with the Cho and Kreps (1987) refinement D1.

to do) rather than going to trial (where there is a chance of conviction and a chance of acquittal, with corresponding informal sanctions), which is the resulting term now on the right (i.e., $S_c(1 - F_G) + k^D$). If there were no informal sanctions for D, then Condition 1 would be satisfied automatically.

Thus, positive informal sanctions on D constrain P's ability to settle cases via plea bargain. Were r^D to violate Condition 1, the feasible set of plea offers that P could optimize over would be empty, resulting in evisceration of plea bargaining. This means that beliefs by P (and Θ) would be given by the prior, and according to MR2, P would always go to trial.

The second issue of concern (item (2) in the first paragraph of this sub-section) is that the informal sanctions may result in P defecting from her part of the equilibrium by making a larger plea offer that provokes both types to reject. This could occur if, despite the presence of informal sanctions in P's expected payoff from trial, $S_b(\rho^D_G) = \pi^D_T(G; \rho^D_G) - r^D$ was less than what P could obtain by driving those D's that would have otherwise settled to trial. In the Technical Appendix we derive Condition 2 as the restriction that eliminates this incentive for defection by P.[25]

Condition 2 (no defection). For P to find it preferable to settle with a D of type G rather than provoking trial (holding Θ's beliefs fixed at the equilibrium ρ^D_G), it must be that:

$$r^D \leq [k^P + k^D + r^P_I\mu(I \mid c)(1 - F_G) + r^P_G\mu(G \mid a)F_G]/[1 - \mu(G \mid c)(1 - F_G) - \mu(G \mid a)F_G].$$

The point of displaying this condition is that while r^D again appears on the left (and the denominators of the expressions in the two Conditions are the same), the informal sanctions on P, at rates r^P_I and r^P_G, contribute to the magnitude of the right-hand-side, and therefore alsso affect the ability to conclude a successful plea bargain.

Finally, P could also defect by dropping all cases; again, this would not change the

[25] Condition 2 is not necessary in the (1, 1, 1) equilibrium (i.e., when all types reject P's offer).

observers' beliefs, but we need to verify that P prefers the hypothesized equilibrium outcome to what she would get by defecting to dropping all cases. Condition 1, however, is sufficient to imply this preference. To see why, notice that in the hypothesized equilibrium $(S_b(\rho_G^D), \rho_G^D, 1, 1)$, P's payoff is:

$$(1 - \rho_G^D)(1 - \lambda)[\pi_T^D(G; \rho_G^D) - r^D] + (\rho_G^D(1 - \lambda) + \lambda)\pi_T^P(\rho_G^D).$$

It is straightforward to show that P is indifferent between trying and dropping the case for $\rho_G^D = \rho_G^{D0}$ (and strictly prefers trying to dropping for $\rho_G^D > \rho_G^{D0}$; see the Technical Appendix). Then Condition 1 implies that the settlement offer $S_b(\rho_G^D) = \pi_T^D(G; \rho_G^D) - r^D$ is non-negative, whereas P's payoff from dropping the case is $- r_G^P\mu(G \mid d)$, which is strictly negative. Thus, P strictly prefers the outcome involving some plea bargaining to defecting to dropping all cases.

Since P prefers to settle via plea bargain rather than going to trial against type G when Condition 2 holds, why can't P offer a slightly lower plea offer than $S_b(\rho_G^D) = \pi_T^D(G; \rho_G^D) - r^D$ and induce type G to accept with probability 1? The reason is that P needs to maintain her own incentives to go to trial following rejection. If P were to offer a slightly lower plea offer designed to induce type G to accept for sure, then *G should expect the case to be dropped following rejection* (since P's subsequent beliefs should be that every rejection is coming from a D of type I). This will cause type G to reject this offer, despite its apparent allure. Consequently, P cannot gain by deviating to a lower plea offer.[26]

Selecting an Equilibrium

Proposition 1 characterizes the nature of the equilibria for the game, but we are still left with a continuum of equilibria. We now propose a basis to select a unique member of that family, namely

[26] For the equilibrium at $\rho_G^D = \rho_G^{D0}$, there is actually a continuation equilibrium following $S_b < S_b(\rho_G^D)$ wherein type G mixes between accepting and rejecting the plea offer with probability ρ_G^{D0} and P mixes between taking the case to trial and dropping it; see the Appendix for details.

$(S_b(\rho_G^{D0}),\rho_G^{D0}, 1, 1)$, which is the equilibrium with the highest fraction of G-types that choose to accept the plea bargain.

Notice that Θ's beliefs punish I-types by lumping them in with G-types. For example, $r^D\mu(G \mid c)$ is the sanction for a D that is convicted, whether he is truly guilty or innocent; if Θ knew that D was a wrongly-convicted I, then one would expect the sanction to be (at worst) zero. Θ misclassifies an I as a G, based on observing a conviction, with probability $\lambda(1 - F_I)$, so the misclassification cost is $\lambda(1 - F_I)r^D\mu(G \mid c)$. Similarly, if P obtains a conviction against D, then she will suffer an informal sanction based on Θ's beliefs that the convicted D might be innocent, in the amount of $r_I^P\mu(I \mid c)$. But if Θ knew that D was a wrongly-convicted I, then the appropriate informal sanction for P would be r_I^P. Thus the misclassification cost is $\lambda(1 - F_I)[r_I^P - r_I^P\mu(I \mid c)]$.

In the Appendix we assume that these costs are additive and we provide an overall "regret" measure, denoted as $R(\rho_G^D)$, which is shown to reduce to the following expression:

$$R(\rho_G^D) = (r^D + r_I^P)\{\lambda(1 - F_I)\mu(G \mid c) + \rho_G^D(1 - \lambda)(1 - F_G)\mu(I \mid c)\}$$
$$+ (r^D + r_G^P)\{\lambda F_I\mu(G \mid a) + \rho_G^D(1 - \lambda)F_G\mu(I \mid a)\}. \qquad (13)$$

We further show that $R(\rho_G^D)$ is increasing on $[\rho_G^{D0}, 1]$. That is, if we assume that outside observers, while not bearing any direct costs for imposing informal sanctions, have a preference to have a legal system that allows them to make the smallest level of classification errors, then *to minimize the regret observers experience due to erroneously-imposed sanctions*, the observers should adopt the conjecture ρ_G^{D0}, and the associated beliefs. In Proposition 2 we adopt this notion to select the specific equilibrium $(S_b(\rho_G^{D0}), \rho_G^{D0}, 1, 1)$.

Proposition 2. If r^D is not "too large" and if the observers adopt the regret-minimizing conjecture as to a G-type's strategy to accept or reject a plea bargain, then the unique equilibrium for the game is $(S_b(\rho_G^{D0}), \rho_G^{D0}, 1, 1)$. In equilibrium, P makes the plea bargain offer $S_b(\rho_G^{D0})$, G-types reject the offer with probability $\rho_G^{D0} < 1$, I -types always reject the offer, and P always takes all Ds that reject the plea offer to trial.

Notice that the notion of preferring a lower value of ρ_G^D is also consistent with a "veil of ignorance" argument for a Θ who realizes they might become a D some day. Higher values of ρ_G^D than ρ_G^{D0} mean lower payoffs for both G-types and I-types, so while a Θ does not incur a direct cost from applying sanctions (and presumably does not derive utility from applying such sanctions), behind a veil of ignorance as to whether or not Θ might become a D, any positive probability associated with that possibility means that Θ should prefer ρ_G^{D0} to any higher value of ρ_G^D.[27]

A special subcase of our model involves eliminating all the informal sanctions, so that $r^D = r_I^P = r_G^P = 0$. Again, the G-type mixes because P is not committed to going to trial, so some fraction of the G-types must reject the offer in order that P not defect to dropping cases (which is due to MR1). In this case Θ's beliefs do not affect D or P. Conditions 1 and 2 now always hold, so there is always a plea bargain, $S_b(\rho_G^{D0})$, that P makes in equilibrium. Thus, we can see that it is the informal sanctions that can restrict or eliminate plea bargaining.

Comparative Statics

Table 1 provides a summary of effects of changes in the informal sanctions on the equilibrium likelihood of plea bargaining failure (ρ_G^{D0}), the size of P's equilibrium offer (S_b), on Θ's beliefs about whether dispositions of cases at trial are "justly" well-correlated with the types of D

[27] Notice also that the equilibrium in Proposition 1 wherein $\rho_G^D = 1$ provides the same payoffs as if there were no plea bargaining. Thus, since the selected equilibrium involves $\rho_G^D = \rho_G^{D0} < 1$, then both types of defendant and the outside observer all prefer that plea bargaining be possible. This reflects the externality that even though all I-types reject the plea offer and go to trial, the fact that some G-types accept the offer reduces the guilty types in the pool of defendants choosing trial, thereby raising the equilibrium belief of innocence for defendants who choose to reject the plea offer.

(that is, $\mu(G \mid a)$ and $\mu(I \mid c)$),[28] and about the right-hand-sides of Conditions 1 and 2.

Table 1: Primary Comparative Statics

	ρ_G^{D0}	$S_b(\rho_G^{D0})$	$\mu(G \mid a)$	$\mu(I \mid c)$	Cond'n 1	Cond'n 2
r^D	0	–	0	0	0	0
r_I^P	+	+	+	–	+	+
r_G^P	–	–	–	+	–	**?**

Thus, for example, increases in r^D have no effect on the likelihood of bargaining failure, the beliefs, or the right-hand-sides of the two conditions (trivially, r^D affects the left-hand-sides of the conditions). Increases in either of r_I^P or r_G^P have opposite effects on the first five columns of the Table, and all of this springs from their effect on ρ_G^{D0}, which by direct computation is (see the Appendix):

$$\rho_G^{D0} = - \lambda[(S_c - r_I^P)(1 - F_I) - k^P]/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P]. \tag{14}$$

For example, consider an increase in r_I^P; it enters ρ_G^{D0} via the numerator, so according to equation (14) increasing r_I^P means that the computed value of ρ_G^{D0} should (in equilibrium) rise. The intuition for this rise in ρ_G^{D0} is that increasing r_I^P makes trial less appealing to P, since it is the errors arising from trial that can result in an I-type being convicted. This undermines P's incentives to take Ds who reject the offer to trial, so to maintain those incentives, more G-types must be in the mix, thereby requiring that more reject the offer. More G-types in the mix allows the equilibrium plea bargain to rise. Furthermore, higher values of r_I^P, which causes a greater fraction of G-types to reject the offer, means that the pool of Ds at trial is richer in G-types, meaning that both conviction and acquittal are more likely to be associated with a G-type. This in turn means that conviction is less

[28] Parameter adjustments that increase these particular beliefs would seem to increase the degree of perceived injustice or unfairness of the legal system.

likely to be associated with an I-type (since $\mu(I \mid c) = 1 - \mu(G \mid c)$).

Using the results from Table 1 and some algebra, an increase in r_I^P increases the right-hand-side of Condition 1, and with yet more effort, an increase in r_I^P increases the right-hand-side of Condition 2. Further, an increase in r_G^P decreases the right-hand-side of Condition 1, so greater social opprobrium towards both defendants and towards P's who might be viewed as "soft on crime" (that is, enabling the guilty to escape justice) means that the options for P to successfully conclude a plea bargain resolution of a criminal case go down. Finally, and unfortunately, the effect of an increase in r_G^P on the right-hand-side of Condition 2 is not clear.[29]

Police Arrest and Trial Process Effectiveness

Earlier we reflected on police arrest effectiveness as involving λ being small: the intake process into the legal system is effective if the likelihood that the police arrested an I-type is (relatively) small. From equation (14) above reductions in λ lead to reductions in ρ_G^{D0}, meaning that improvements in police arrest effectiveness increases the likelihood of plea bargaining success, which reduces the expenditure of court costs, since fewer cases go to trial. Moreover, since ρ_G^{D0} enters the beliefs this means that, in equilibrium, the beliefs $\mu(G \mid a)$ and $\mu(I \mid c)$ are independent of λ, suggesting that the intervening bargaining game between arrest and trial eliminates the impact of the prior λ on the posterior assessments about the results of trial that Θ makes.

We are also interested in the effectiveness of the trial process at properly convicting the guilty and acquitting the innocent. If trials were perfect discriminators of guilt or innocence, we would have $F_I = 1$ (all innocent Ds are acquitted) and $F_G = 0$ (all guilty Ds are convicted). While this thought experiment does not seem possible to achieve, it does inform us about what we want

[29] Its effect appears to be complexly-related to the other parameter values as well as the characteristics of the F-distribution.

investment in trial resources to do (such investments might involve better procedures for obtaining and vetting evidence, or improved procedures for the trial itself).

Let z be a level of investment in trial resources, and now extend our earlier notation $F(\gamma_c \mid t)$ to incorporate investment z:

$$F(\gamma_c \mid t, z),\ t = I, G,$$

denotes the probability that type t's evidence draw yields an acquittal if the investment level is z and the evidentiary standard for conviction is γ_c. Following our earlier example of a perfect trial, this means that we want to require that:

$$F_{Iz} \equiv \partial F(\gamma_c \mid I, z)/\partial z > 0 \text{ and } F_{Gz} \equiv \partial F(\gamma_c \mid G, z)/\partial z < 0.$$

Such an investment would increase trial effectiveness (we abstract from concerns about the cost of such investments). Some investments may only affect F_I or F_G, while some might affect both.

Again, returning to equation (14), an increase in z that only affects F_G thereby increases the only the denominator of ρ_G^{D0}, meaning that such an investment increases the likelihood of plea-bargaining success. The stand-alone effect of z via F_I is more complicated. This is because of the relationship between S_c and r_I^P. Recall that MR1 required that $(S_c - r_I^P)(1 - F_I) - k^P < 0$: if P (and Θ) know that D was of type I, P should prefer to drop the case rather than take the case to trial. However, this does not determine the sign of $S_c - r_I^P$. If this term is positive, then from P's perspective, her payoff (abstracting from the cost of trial) from a conviction of an innocent D exceeds the informal sanction rate, while if this term is negative, it means that the informal sanction rate outweighs the utility P gains from obtaining the formal sanction.

Now we can turn to the question of an investment that improves the likelihood of acquitting an innocent D. If the informal sanction rate is small ("small r_I^P," meaning $S_c - r_I^P > 0$), then an increase in z leads to an increase in F_I, which leads to an increase in ρ_G^{D0}, which means less settlement, and higher plea offers as well (since $S_b(\rho_G^{D0})$ also rises in equilibrium). If the informal sanction rate is large ("large r_I^P," meaning $S_c - r_I^P < 0$), then an increase in z leads to an increase in F_I, which leads to a decrease in ρ_G^{D0}, which means more settlement, and lower plea offers as well (since $S_b(\rho_G^{D0})$ also falls in equilibrium). Now putting this together with the effects on F_G, we see that the effect of an increase in z in the large-r_I^P case definitely leads to a reduction in ρ_G^{D0} (and in $S_b(\rho_G^{D0})$), as the numerator of equation (14) is falling while the denominator of equation (14) is increasing. Here, trial is becoming clearly more effective at separating I-types and G-types and providing them with the corresponding dispositions of a and c, respectively: trial is a better tool for <u>not</u> convicting the innocent and <u>for</u> convicting the guilty. Sadly, this clarity is not present in the small-r_I^P case when investment affects both F_I and F_G because it is not possible to sign the effect of z on ρ_G^{D0}, since both the numerator and the denominator of equation (14) are increasing.

4. Refining the Jury's Assessment: The Scottish Verdict

For almost 300 years, Scotland has used a three-outcome verdict for criminal juries; a defendant is found not guilty, or not proven, or guilty, with no formal sanction attaching to the the first two outcomes.[30] Such a refinement of the jury's assessment of a defendant's guilt or innocence should provide more information to the outside observers to employ in applying informal sanctions. Does it and, if so, what else does it do?

To address this extension, we strengthen an earlier assumption about the pair of distributions F(e | t), t = I, G. We now assume that the strict monotone likelihood ratio property (SMLRP) holds:

---

[30] See Duff (1999) for an extensive discussion of the history of the development of this institution. Bray (2005) indicates that the same three-outcome verdict was used in the 1807 trial of Aaron Burr for treason.

SMLRP: $f(e \mid G)/f(e \mid I)$ is strictly increasing in e, for e in (0, 1). (15)

As discussed in the Technical Appendix, this assumption implies: 1) $F(e \mid I) > F(e \mid G)$ (strict stochastic dominance by G); 2) $f(e \mid G)/(1 - F(e \mid G)) > f(e \mid I)/(1 - F(e \mid I))$ (strict hazard rate dominance by G); and 3) $f(e \mid G)/F(e \mid G) > f(e \mid I)/F(e \mid I)$ (strict reverse hazard rate dominance by G). We represent the three-outcome verdict by the triple {ng, np, g}, with the obvious interpretation, and we assume that $\gamma_g \equiv \gamma_c$ (that is, the same evidentiary standard for a conviction under the previous two-outcome verdict is used to find a defendant "guilty" under the three-outcome verdict). Further, let γ_{ng} be the cutoff for not guilty versus not proven, where $0 < \gamma_{ng} < \gamma_g$. Thus, more formally, we extend the previous notation so that $F_t(\gamma_g) \equiv \Pr\{e \leq \gamma_g \mid t\}$ and $F_t(\gamma_{ng}) \equiv \Pr\{e \leq \gamma_{ng} \mid t\}$, for t = I, G.

In the Technical Appendix we show that:

1) For any non-zero vector of strategies by D, (ρ^D_G, ρ^D_I), Θ's beliefs as to D's likelihood of actually being of type G, having observed one of the mutually-exclusive outcomes ng, np, or g, satisfies:

$$\mu(G \mid ng) < \mu(G \mid ng) < \mu(G \mid g); \tag{16}$$

and 2) that I's expected loss from proceeding to trial $(\pi^D_T(I))$ is strictly lower than G's expected loss from proceeding to trial $(\pi^D_T(G))$, where:

$$\pi^D_T(t) = S_c(1 - F_t(\gamma_g)) + k^D + r^D\mu(G \mid g)(1 - F_t(\gamma_g)) + r^D\mu(G \mid ng)F_t(\gamma_{ng})$$
$$+ r^D\mu(G \mid np)(F_t(\gamma_g) - F_t(\gamma_{ng}))\}, t \in \{I, G\}. \tag{17}$$

As earlier, the ordering of payoffs indicated above means that Proposition 1 applies to the modified game, so that the family of equilibria again involve I-types always rejecting the plea offer and P always taking any D who rejects a plea offer to trial, while G-types mix between accepting the plea offer and rejecting it with probability ρ_G^D. Similarly, P's expected payoff from trial (π_T^P) can be extended to allow for the three outcomes but, as shown in the Technical Appendix, this function is independent of γ_{ng}. This means that since ρ_G^D makes P indifferent between dropping and going to trial, then the equilibrium values for ρ_G^D are the same as in the two-outcome verdict regime. Thus, $\rho_G^D \in [\rho_G^{D0}, 1]$, as before. This occurs because P's reduced-form expected payoffs from trial simply reflect whether D is found guilty or is acquitted. Furtheremore, as shown in the Technical Appendix, Conditions 1 and 2 now hold for a larger range of the parameter r^D, so the use of the three-outcome verdict means that informal sanctions are less likely to interfere with plea bargaining. Proposition 2 carries over to the modified game as well: the selected equilibrium value for ρ_G^D is the same as earlier, ρ_G^{D0}.

The change in the number of possible verdict outcomes affects Θ and D. The difference between the regret functions for the two-outcome verdict and the three outcome verdict is that now the term $\mu(G \mid np)(F_t(\gamma_g) - F_t(\gamma_{ng})) + \mu(G \mid ng)F_t(\gamma_{ng})$ replaces $\mu(G \mid a)F_t(\gamma_g)$ in the computation. This change also appears in D's expected payoff from the three-outcome trial verdict. Using the assumption SMLRP, we show in the Technical Appendix that the following result holds:

$$\mu(G \mid np)(F_G(\gamma_g) - F_G(\gamma_{ng})) + \mu(G \mid ng)F_G(\gamma_{ng}) > \mu(G \mid a)F_G(\gamma_g), \tag{18a}$$

and $$\mu(G \mid np)(F_I(\gamma_g) - F_I(\gamma_{ng})) + \mu(G \mid ng)F_I(\gamma_{ng}) > \mu(G \mid a)F_I(\gamma_g). \tag{18b}$$

The implication of inequalities (18a) and (18b) is summarized in Proposition 3.

Proposition 3: The expected payoff for a D of type G is higher under the three-outcome verdict than under the two-outcome verdict. The expected payoff for a D of type I is lower under the three-outcome verdict than under the two-outcome verdict.

This means that while a G-type still rejects the equilibrium plea offer at the same rate as before, ρ_G^{D0}, the equilibrium offer itself is larger, since $S_b(\rho_G^{D0}) = \pi_T^D(G; \rho_G^{D0}) - r^D$, but the expected payoff for the G-type has gotten worse (a larger loss). Thus, P's overall payoff increases (since the plea bargains are tougher and are accepted at the same rate, and P's trial payoff is unchanged). Finally, it is also straightforward to show that Θ's expected regret is lower in the three-outcome verdict regime. We take this to mean that, overall, use of the Scottish verdict would enhance justice: the I-types lose less, G-types lose more, and Θs impose fewer erroneous informal sanctions.

5. Summary and Further Discussion

Our model considers the strategic interaction between a prosecutor and a defendant when informal sanctions by third parties can be imposed on both the defendant and the prosecutor. These sanctions affect the feasibility of plea bargaining, as well as the level of the bargain offered and the frequency of bargaining success. The model follows the action from choosing a plea offer up through trial, allowing for dropping of cases, thereby not relying on prosecutors being able to pre-commit to taking defendants who reject offers to trial. The defendant's private information concerns his guilt or innocence of the crime for which he was arrested; this underlying state of the world affects the evidence that is presented at trial. Third parties form beliefs about the defendants who are processed through the system, allowing for outcomes wherein a plea is accepted, or a case is dropped, or a defendant goes to trial and is convicted or acquitted. Significantly, while the third parties can observe the disposition of the cases, they cannot observe plea offers or evidence generation; of course they are rational and can construct the equilibrium of the game, but the errors in the legal process (as well as hidden information) means that they will misclassify defendants and thereby erroneously impose sanctions on both defendants and prosecutors.

We show that there is a unique family of equilibria and, if third parties prefer a legal system with minimal regret arising from classification errors, a unique equilibrium within this family, wherein the guilty defendant accepts the prosecutor's proposed plea offer with positive (but fractional) probability, the innocent defendant rejects the proposed offer, and the prosecutor chooses to take all defendants who reject the offer to trial. The plea offer and the decisions by each agent are all a function of the informal sanctions.

We find that informal sanctions may act to constrain the set of possible plea offers the prosecutor can make, but informal sanctions on the prosecutor can work in opposite directions depending upon whether the concern is for convicting the innocent or letting the guilty escape justice. If the informal sanction rate on defendants is high enough (if society is sufficiently condemning of those agents who are arrested), then plea bargaining is eviscerated. High potential sanctions on prosecutors can also lead to distortions, with high sanctioning of prosecutors for likely conviction of innocents leading to tougher plea offers, and more guilty defendants rejecting the plea offer and going to trial, which increases the association between any trial outcome (conviction or acquittal) and guilt. Alternatively, high potential informal sanctions of prosecutors for perceived release of the guilty ("soft on crime") leads to greater use of plea bargaining, as offers are made less tough (acceptance of an offer by a defendant self-labels them as guilty, so prosecutors do not suffer any informal sanctions following a plea bargain).

We further consider the importance of police effectiveness (that police have a solid basis for an arrest in the first place, thereby reducing the likelihood that innocents are swept up into the process) as well as trial effectiveness (the degree to which evidence acts to correctly classify defendants as to their guilt). Increased police effectiveness leads to greater use of plea bargaining to resolve offenses, and thereby saves trial costs. Increased investment in the trial process, when informal sanctions on prosecutors for possibly "railroading" innocents (that is, r_I^P) is sufficiently high, has a clear result of enhancing the effectiveness of plea bargaining; results are less clear if the informal sanctions are less important to the prosecutor than her utility for the formal sanction associated with conviction at trial.

Finally, we examined the effect of extending the analysis to the "Scottish" (three-outcome) verdict, wherein juries find a defendant not guilty, not proven, or guilty. We showed that the propositions proved earlier carry over to the new regime. In particular, the same equilibrium obtains, including the same likelihood of bargaining success, with the only modification being that the equilibrium settlement offer is higher. This reflects the result that, at trial, guilty defendants do worse under this scheme while innocent defendants do better. Moreover, outside observers impose informal sanctions with a lower total regret associated with erroneous application. Overall, the Scottish verdict leads to an increase in justice.

Further Discussion Concerning Maintained Assumptions and Possible Extensions

An important assumption made in the model is that all agent payoffs are linear in sanctions, both formal and informal (see the payoffs in Section 2). This provided an important result: innocent defendants expect a smaller loss from proceeding to trial than guilty ones (Remark 1). Using this result contributes to the further result that we can (at least partly) separate the guilty from the innocent via a plea offer, with (in equilibrium) all innocent defendants rejecting the offer, and some guilty defendants also rejecting it.

It is quite possible that defendants may have a nonlinear response to risk, and the implications for our equilibria may be significant.[31] If the G-types are less risk-averse (or possibly even risk-taking) in comparison with I-types, then Remark 1 may not hold. This, in turn, would alter the type of equilibrium for the game. For example, the payoffs might reverse their ordering and we might have guilty defendants electing trial and innocent defendants being willing to accept a plea offer. Or, more interestingly, neither payoff would dominate for all possible distributions of evidence, leading to an equilibrium with both types randomizing between accepting and rejecting an offer, and the prosecutor also randomizing on trial versus dropping. We view this as a worthwhile, but very complex, extension.[32]

A second direction of extension would be to provide a more complete model of the trial, especially the generation of evidence. In the current analysis, the only conditioning variable for the F-distribution is the underlying guilt or innocence of the defendant. If the prosecutor could make an investment in evidence production, possibly based on what she inferred from the plea bargaining outcome, that might affect both the trial and the plea bargaining process itself (especially if plea bargaining was modeled as – say – an alternating-offer process, rather than a take-it-or-leave-it offer as we do above). Of course, allowing the prosecutor to make such investments suggests allowing the defendants to do this as well, and that suggests a more sophisticated model of the jury's decision process might also be warranted.[33]

Third, one might use the model to consider possible social policies to rectify the inaccuracies

[31] Becker (1968) indicates that offenders (G-types) may be risk-takers, while law-abiding citizens may be risk-avoiders.

[32] If G-types are more likely to have experience with the legal process than I-types, then it is possible that G-types can better anticipate the relevant details of the distribution associated with evidence, F, while I-types might be unsure about the F-distribution that would apply. That is, perhaps I-types are ambiguity averse, and this, too, would affect payoffs. For a recent application of ambiguity aversion to civil suits, see Franzoni (2014).

[33] It may also be worth noting that police effectiveness might affect the distributions of evidence as well, which might cause the observers' equilibrium posterior beliefs to be influenced by the prior.

inherent in the use of informal sanctions. Should society compensate those who are acquitted for their trial costs? For the inevitable informal sanction that they face, even though they are acquitted? How would the presence of such a subsidy impact plea bargaining? Would greater transparency of the legal process further help reduce misapplication of informal sanctions? We have the jury following a simple rule (i.e., not drawing an inference from the level of the evidence, e, as to the underlying distribution – and therefore type of D). What if outside observers could see the evidence and draw inferences from it? Both of these suggestions (subsidies and transparency) would involve substantial modifications to the game, but could lead to interesting implications.

Finally, another possible direction of extension concerns how informal sanctions affect deterrence and the application of both formal and informal sanctions.[34] This extension necessitates expanding the model so as to allow for a socially optimal choice of the statutory sanction for convictions, S_c. The presence of informal sanctions should affect the socially optimal level of the formal sanctions, though the direction is unclear. Informal sanctions on the defendant suggest that the statutory sanction would be lower (assuming that the existing level ignored the possibility of informal sanctions), since it is the total level that would influence the person committing the crime, and incarceration implies costs due to maintaining facilities as well as any productivity losses. Thus, this might lead to lower plea offers, but lower statutory sanctions, coupled with informal sanctions on prosecutors, could lead to reduced incentives for prosecutors, which may be undesirable. Whether improved deterrence would have any effect on classification error in the application of the informal sanctions would also be worth evaluating.

---

[34] See Reinganum (1993) for an example of a model with criminals choosing whether to commit a crime, police potentially detecting a crime committed, and plea bargaining with a prosecutor if arrested.

References

Associated Press, "Day in Jail for Ex-Duke Prosecutor," *New York Times*, September 1, 2007. Online at http://www.nytimes.com/2007/09/01/us/01nifong.html; accessed on May 20, 2014.

Baker, Scott and Mezzetti, Claudio, "Prosecutorial Resources, Plea Bargaining, and the Decision to Go to Trial," *Journal of Law, Economics, and Organization*, 17(1), 2001, pp. 149-167.

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 1968, pp. 169-217.

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

Boylan, Richard T. and Long, Cheryl X., "Salaries, Plea Rates, and the Career Objectives of Federal Prosecutors," *Journal of Law and Economics*, 48(2), 2005, pp. 627-652.

Boylan, Richard T., "What do Prosecutors Maximize? Evidence from the Careers of U.S. Attorneys," *American Law and Economics Review*, 7(2), 2005, pp. 379-402.

Bray, Samuel, "Not Proven: Introducing a Third Verdict," *The University of Chicago Law Review*, 72(4), 2005, pp.1299-1329.

Burke, Mary Kate; Hopper, Jessica; Francis, Enjoli; and Effron, Lauren, "Casey Anthony Juror: 'Sick to Our Stomachs' Over Not Guilty Verdict," ABC News, July 6, 2011. Online at: http://abcnews.go.com/US/casey_anthony_trial/casey-anthony-juror-jury-sick-stomach-guilty-verdict/story?id=14005609

Duff, Peter, "The Scottish Criminal Jury: A Very Peculiar Institution," *Law and Contemporary Problems*, 62(2), 1999, pp. 173-201.

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

Franzoni, Luigi A., "Liability Law and Uncertainty Spreading," Working Paper, Department of Economics, University of Bologna, February, 2014.

Grossman, Gene M. and Katz, Michael L., "Plea Bargaining and Social Welfare, *The American Economic Review*, 73(4), 1983, pp. 749-757.

Iacobucci, Edward M., "On the Interaction between Legal and Reputation Sanctions," *Journal of Legal Studies*, 43(1), 2014, pp. 189-207.

Landes, William M., "An Economic Analysis of the Courts," *Journal of Law and Economics*, 14(1), 1971, pp. 61-108.

Nalebuff, Barry, "Credible Pretrial Negotiation, *The RAND Journal of Economics*, 18(2), pp. 198-210.

NC State Bar v. Michael B. Nifong, Disciplinary Hearing Commission, NC State Bar, 06 DHC 35. Online at http://www.ncbar.com/discipline/printorder.asp?id=505

Reinganum, Jennifer F., "Plea Bargaining and Prosecutorial Discretion," *The American Economic Review*, 78(4), 1988, pp. 713-728.

Reinganum, Jennifer F., "The Law Enforcement Process and Criminal Choice," *International Review of Law and Economics*, 13(2), 1993, pp. 115-134.

Segal, David, "Mugged by a Mug Shot Online," *New York Times*, October 5, 2013. Online at http://www.nytimes.com/2013/10/06/business/mugged-by-a-mug-shot-online.html?pagew; accessed on May 20, 2014.

Appendix

Characterizing Equilibria

The only candidates for an equilibrium involve $\rho_I^D = 1$ (I-types always reject the plea offer) and $\rho_G^D \in (0, 1]$ (G-types always reject the plea offer with positive probability); see the Technical Appendix wherein the other candidates are ruled out. P may also mix between taking the case to trial and dropping it following a rejection.

The timing of the game is such that each type of D chooses to accept or reject the plea offer, taking as given the likelihood that P takes the case to trial following rejection; and P chooses to take the case to trial or drop it, given her beliefs about the posterior probability that D is of type G, given rejection. Both of these decisions are taken following P's choice of plea offer, S^P, so both parties must take this offer as given at subsequent decision nodes.

We first characterize the equilibrium in the continuation game, given S_b, allowing for mixed strategies for both P (ρ^P) and the D of type G (ρ_G^D). Since the observers' beliefs will depend on their conjectured value for ρ_G^D, we will augment the notation for the observers' beliefs to reflect these conjectures, $\rho_G^{D\Theta}$. Other functions that also depend on these conjectures through the observers' beliefs will be similarly augmented.

Suppose that observers conjecture that the D of type G rejects the plea offer with probability $\rho_G^{D\Theta}$. Then $\mu(G \mid c; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)(1 - F_G)/[\rho_G^{D\Theta}(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]$; $\mu(G \mid a; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)F_G/[\rho_G^{D\Theta}(1 - \lambda)F_G + \lambda F_I]$; $\mu(G \mid d; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \lambda]$; and $\mu(G \mid b; \rho_G^{D\Theta}) = 1$. Moreover, suppose that the D of type G anticipates these beliefs, and also expects that P will take the case to trial following rejection with probability ρ^P. Then type G will be indifferent, and hence willing to mix, between accepting and rejecting the offer S_b, if $\pi_R^D(G; \rho_G^{D\Theta}) = \rho^P\pi_T^D(G; \rho_G^{D\Theta}) + (1 - \rho^P)\pi_d^D(\rho_G^{D\Theta}) = \pi_b^D(\rho_G^{D\Theta})$. Substitution and simplification yields the value of ρ^P that renders G indifferent:

$$\rho^P(S_b; \rho_G^{D\Theta}) = \frac{\{S_b + r^D(1 - \mu(G \mid d; \rho_G^{D\Theta}))\}}{\{S_c(1 - F_G) + k^D + r^D[\mu(G \mid c; \rho_G^{D\Theta})(1 - F_G) + \mu(G \mid a; \rho_G^{D\Theta})F_G - \mu(G \mid d; \rho_G^{D\Theta})]\}}.$$

The numerator of the expression $\rho^P(S_b; \rho_G^{D\Theta})$, which is the difference between type G's payoff from accepting the plea offer versus having his case dropped, is clearly positive, meaning that D would prefer to have his case dropped than to accept a plea offer. The denominator of the expression $\rho^P(S_b; \rho_G^{D\Theta})$ is the difference between type G's payoff from trial versus having his case dropped. This denominator is also positive (see Remark 3 in the Technical Appendix), which implies that type G would prefer that P drop the case against him rather than take it to trial.

Since the observers' beliefs are based on their conjectures $\rho_G^{D\Theta}$ and the case disposition, and NOT on S_b, which they do not observe, the expression $\rho^P(S_b; \rho_G^{D\Theta})$ is an increasing function of S_b. That is, when S_b is higher, P must take the case to trial following rejection with a higher probability in order to make the D of type G indifferent about accepting or rejecting S_b. Notice that even a plea offer of $S_b = 0$ requires a positive probability of trial following a rejection in order to induce the D of type G to be willing to accept it; this is because acceptance of a plea offer comes with a sure informal sanction of r^D (as only a truly guilty D is expected to accept the plea).

Now consider P's decision about trying versus dropping the case. Again suppose that observers – and P – both conjecture that type G rejects the plea offer with probability $\rho_G^{D\Theta}$ in this candidate for equilibrium; thus $\nu(G \mid R; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \lambda]$. Since these conjectures must be the same (and correct) in equilibrium, it is valid to equate them at this point in order to identify what common beliefs for P and Θ will make P indifferent, and hence willing to mix, between trying and dropping the case following a rejection. P will be indifferent between these two options if $\pi_T^P(\rho_G^{D\Theta}) = \pi_d^P(\rho_G^{D\Theta})$; that is, if:

$$\nu(G \mid R; \rho_G^{D\Theta})\{S_c(1 - F_G) - k^P - r_I^P\mu(I \mid c; \rho_G^{D\Theta})(1 - F_G) - r_G^P\mu(G \mid a; \rho_G^{D\Theta})F_G\}$$

$$+ \nu(I \mid R; \rho_G^{D\Theta})\{S_c(1 - F_I) - k^P - r_I^P\mu(I \mid c; \rho_G^{D\Theta})(1 - F_I) - r_G^P\mu(G \mid a; \rho_G^{D\Theta})F_I\} = - r_G^P\mu(G \mid d; \rho_G^{D\Theta}).$$

Substituting for the beliefs and solving for the value of $\rho_G^{D\Theta}$ that generates this equality (see the Technical Appendix for details) yields:

$$\rho_G^{D0} = - \lambda[(S_c - r_I^P)(1 - F_I) - k^P]/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P],$$

where the numerator is positive by MR1; MR2 implies that the denominator is positive and the ratio is a fraction. For any $\rho_G^{D\Theta} > \rho_G^{D0}$, P will strictly prefer to take the case to trial following a rejection, and for any $\rho_G^{D\Theta} < \rho_G^{D0}$, P will strictly prefer to drop the case following a rejection.

To summarize, type G is willing to mix between accepting and rejecting the plea offer S_b if he anticipates that the observers' beliefs are $\rho_G^{D\Theta} = \rho_G^{D0}$ and he expects that P will take the case to trial following rejection of offer S_b with probability $\rho^P(S_b; \rho_G^{D0})$. P is indifferent between trying and dropping the case if she and the observers believe that type G rejects the plea offer with probability ρ_G^{D0}. Thus, the mixed-strategy continuation equilibrium, given S_b, is $(\rho_G^{D0}, \rho^P(S_{b;} \rho_G^{D0}))$.

We can now move back to the decision node at which P chooses the plea offer S_b, anticipating that it will be following by the mixed-strategy equilibrium $(\rho_G^{D0}, \rho^P(S_b; \rho_G^{D0}))$ in the continuation game. P's payoff from making the plea offer S_b is:

$$(1 - \rho_G^{D0})(1 - \lambda)S_b + (\rho_G^{D0}(1 - \lambda) + \lambda)[\rho^P(S_{b;} \rho_G^{D0})\pi_T^P(\rho_G^{D0}) + (1 - \rho^P(S_{b;} \rho_G^{D0}))\pi_d^P(\rho_G^{D0})].$$

The set of feasible S_b values is bounded below by 0 and above by $S_b = \pi_T^D(G; \rho_G^{D0}) - r^D$, where $\pi_T^D(G; \rho_G^{D0})$ is the expression for $\pi_T^D(G)$, evaluated at the beliefs $\mu(G \mid c; \rho_G^{D0}) = \rho_G^{D0}(1 - \lambda)(1 - F_G)/[\rho_G^{D0}(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]$; and $\mu(G \mid a; \rho_G^{D0}) = \rho_G^{D0}(1 - \lambda)F_G/[\rho_G^{D0}(1 - \lambda)F_G + \lambda F_I]$. This is because accepting the plea offer results in a combined sanction of $S_b + r^D$ (since only guilty D's accept the plea offer) and thus any plea offer higher than $\pi_T^D(G; \rho_G^{D0}) - r^D$ will be rejected for sure (rather than with probability ρ_G^{D0}). At this upper bound, the function $\rho^P(S_{b;} \rho_G^{D0})$ just reaches 1. In order to have a non-empty feasible range, we need $\pi_T^D(G; \rho_G^{D0}) - r^D \geq 0$; or, equivalently (note that the denominator of the expression below is positive):

Condition 1. In order for P to be able to induce a D of type G to accept a plea offer, it must be that $r^D \leq [S_c(1 - F_G) + k^D]/[1 - \mu(G \mid c; \rho_G^{D0})(1 - F_G) - \mu(G \mid a; \rho_G^{D0})F_G]$.

Returning to P's payoff as a function of S_b, notice two things. First, since ρ_G^{D0}, which is independent of S_b, renders P indifferent between trying and dropping the case following rejection, the term in square brackets simply equals $\pi_d^P(\rho_G^{D0}) = - r_G^P\mu(G \mid d; \rho_G^{D0})$, where $\mu(G \mid d; \rho_G^{D0}) = \rho_G^{D0}(1 - \lambda)/[\rho_G^{D0}(1 - \lambda) + \lambda]$. Thus, the optimal S_b that supports some plea bargaining is $S_b(\rho_G^{D0}) = \pi_T^D(G; \rho_G^{D0}) - r^D$, the upper limit of the feasible range. This offer is rejected by type G with probability ρ_G^{D0}, and P goes to trial with certainty following a rejection. Note that a D of type I would always reject this plea offer, consistent with the hypothesized form of the equilibrium.

Every plea offer in the feasible set $[0, \pi_T^D(G; \rho_G^{D0}) - r^D]$ is consistent with a mixed-strategy equilibrium in which some D's of type G accept, and others reject, the offer. But – taking the observers' beliefs of ρ_G^{D0} as given – P could make a higher demand that would provoke certain rejection by both D-types. We need to verify that P prefers the hypothesized equilibrium described above to the "defection payoff" she would obtain if all cases went to trial.

In the hypothesized equilibrium, P settles with $(1 - \rho_G^{D0})(1 - \lambda)$ guilty defendants and goes to trial against the rest of the guilty defendants and all of the innocent defendants; if P defects and provokes rejection by all, then she will simply replace the settlement $S_b(\rho_G^{D0}) = \pi_T^D(G; \rho_G^{D0}) - r^D$ with the expected payoff from taking a guilty defendant to trial (holding the observers' beliefs fixed at ρ_G^{D0}). Thus, P prefers (at least weakly) the hypothesized equilibrium to defection as long as:

$$\pi_T^D(G; \rho_G^{D0}) - r^D = S_c(1 - F_G) + k^D + r^D\mu(G \mid c; \rho_G^{D0})(1 - F_G) + r^D\mu(G \mid a; \rho_G^{D0})F_G - r^D$$

$$\geq S_c(1 - F_G) - k^P - r_I^P\mu(I \mid c; \rho_G^{D0})(1 - F_G) - r_G^P\mu(G \mid a; \rho_G^{D0})F_G.$$

Rearranging, we can write this as:

Condition 2. For P to find it preferable to settle with a D of type G rather than provoking a trial (holding the observers' beliefs fixed at ρ_G^{D0}), it must be that:

$$r^D \leq [k^P + k^D + r_I^P\mu(I \mid c; \rho_G^{D0})(1 - F_G) + r_G^P\mu(G \mid a; \rho_G^{D0})F_G]/[1 - \mu(G \mid c; \rho_G^{D0})(1 - F_G) - \mu(G \mid a; \rho_G^{D0})F_G].$$

Finally, P could also defect by dropping all cases; again, this would not change the observers' beliefs, but we need to verify that P prefers the hypothesized equilibrium outcome to what she would get by defecting to dropping all cases. However, Condition 1 is sufficient to imply this preference. To see why, notice that in the hypothesized equilibrium, P's payoff is:

$$(1 - \rho_G^{D0})(1 - \lambda)[\pi_T^D(G; \rho_G^{D0}) - r^D] + (\rho_G^{D0}(1 - \lambda) + \lambda)\pi_T^P(\rho_G^{D0}).$$

We already know that $\pi_T^P(\rho_G^{D0}) = \pi_d^P(\rho_G^{D0})$ by construction. Then Condition 1 implies that the settlement offer $S_c(\rho_G^{D0}) = \pi_T^D(G; \rho_G^{D0}) - r^D$ is non-negative, whereas P's payoff from dropping a case is $- r_G^P\mu(G \mid d; \rho_G^{D0})$, which is strictly negative. Thus, P strictly prefers the outcome involving some plea bargaining to defecting to dropping all cases.

Multiple Equilibria

There can also be equilibria wherein the D of type G rejects the plea bargain with probability $\rho_G^D > \rho_G^{D0}$. The reason for this multiplicity of equilibria is that the observers' beliefs actually drive the equilibrium behavior of P and the D of type G. To see this, suppose that $\rho_G^{D\Theta} = \rho_G^{D1} > \rho_G^{D0}$, and that P and type G anticipate these beliefs. Since the observers' posterior beliefs $\mu(G \mid c; \rho_G^{D\Theta})$ and $\mu(G \mid a; \rho_G^{D\Theta})$ are increasing in $\rho_G^{D\Theta}$, both P and the D of type G expect that D will face harsher informal sanctions following either trial outcome (than they would face at ρ_G^{D0}).

This means that P can demand the higher plea sentence, $S_b(\rho_G^{D1}) = \pi_T^D(G; \rho_G^{D1}) - r^D$, which will make the D of type G indifferent about accepting and rejecting it (if he expects P to take the case to trial following a rejection); thus, type G is willing to randomize and reject the plea bargain with probability equal to ρ_G^{D1}. Since $\rho_G^{D1} > \rho_G^{D0}$, P will take the case to trial with probability 1 following a rejection (if he thinks that the D of type G is using the rejection probability ρ_G^{D1}). Thus, there is an equilibrium at any $\rho_G^D \in [\rho_G^{D0}, 1)$ as long as Conditions 1 and 2 continue to hold at that ρ_G^D.

Since P prefers to settle via plea bargain rather than going to trial against type G when Condition 2 holds, and P strictly prefers to go to trial (rather than drop) following rejection when $\rho_G^D > \rho_G^{D0}$, why can't P offer a slightly lower plea offer than $S_b(\rho_G^D) = \pi_T^D(G; \rho_G^D) - r^D$ and induce type G to accept with probability 1? The reason is that P needs to maintain her own incentives to go to trial following rejection. If P were to offer a slightly lower plea offer designed to induce type G to accept for sure, then *G should expect the case to be dropped following rejection* (since P's subsequent beliefs should be that every rejection is coming from a D of type I). This will cause type G to reject this offer, despite its apparent allure. Consequently, P cannot gain by deviating to a lower plea offer.

Selecting Among Equilibria

The following terms summarize erroneously-imposed informal sanctions. Excessive sanctions imposed on type I defendants following conviction, or acquittal, at trial (ideally, there would be no sanctions):

$$\lambda(1 - F_I)r^D\mu(G \mid c) + \lambda F_I r^D\mu(G \mid a).$$

Insufficient sanctions imposed on type G defendants following conviction, or acquittal, at trial (ideally, the sanction would be r^D):

$$\rho_G^D(1 - \lambda)(1 - F_G)[r^D - r^D\mu(G \mid c)] + \rho_G^D(1 - \lambda)F_G[r^D - r^D\mu(G \mid a)].$$

Prosecutors also suffer erroneously-imposed sanctions. With respect to innocent defendants:

$$\lambda(1 - F_I)[r_I^P - r_I^P\mu(I \mid c)] + \lambda F_I[r_G^P\mu(G \mid a)].$$

Note that the first term reflects the fact that P convicted a type I and ideally would have received the sanction r_I^P, but she only received $r_I^P\mu(I \mid c)$; the second term reflects the fact that an innocent was acquitted, but P was still sanctioned because observers' beliefs admit some probability that the acquittal was an error, for which P is sanctioned (undeservedly).

With respect to guilty defendants:

$$\rho_G^D(1 - \lambda)(1 - F_G)[r_I^P\mu(I \mid c)] + \rho_G^D(1 - \lambda)F_G[r_G^P - r_G^P\mu(G \mid a)].$$

The first term reflects the fact that P convicted a guilty D, but was still sanctioned because observers' beliefs admit some probability that the conviction was an error, for which P is sanctioned (undeservedly); the second term reflects the fact that a guilty D was acquitted, so P ideally would have received the sanction r_G^P but only received $r_G^P\mu(G \mid a)$.

Let $R(\rho_G^D)$ denote the measure of "regret" that observers experience due to these erroneous sanctions. Then (all of the observers' beliefs are evaluated at ρ_G^D):

$$\begin{aligned} R(\rho_G^D) = {} & \lambda(1 - F_I)r^D\mu(G \mid c) + \lambda F_I r^D\mu(G \mid a) + \rho_G^D(1 - \lambda)(1 - F_G)[r^D - r^D\mu(G \mid c)] \\ & + \rho_G^D(1 - \lambda)F_G[r^D - r^D\mu(G \mid a)] + \lambda(1 - F_I)[r_I^P - r_I^P\mu(I \mid c)] + \lambda F_I[r_G^P\mu(G \mid a)] \\ & + \rho_G^D(1 - \lambda)(1 - F_G)[r_I^P\mu(I \mid c)] + \rho_G^D(1 - \lambda)F_G[r_G^P - r_G^P\mu(G \mid a)]. \end{aligned}$$

Collecting terms simplifies the above expression greatly:

$$\begin{aligned} R(\rho_G^D) = {} & (r^D + r_I^P)\{\lambda(1 - F_I)\mu(G \mid c) + \rho_G^D(1 - \lambda)(1 - F_G)\mu(I \mid c)\} \\ & + (r^D + r_G^P)\{\lambda F_I\mu(G \mid a) + \rho_G^D(1 - \lambda)F_G\mu(I \mid a)\}. \end{aligned}$$

Upon recalling the definitions of $\mu(t \mid y)$, and evaluating them at ρ_G^D), it is straightforward (though tedious) to show that both of the terms in curly brackets in the expression $R(\rho_G^D)$ are increasing in ρ_G^D.