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

by Andrew F. Daughety and Jennifer F. Reinganum

Technical Appendix

The basic notation and the building blocks of the payoff functions are given in the text; some of this material is repeated here for easy reference. The D of type t's payoff function from trial is:

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

D's payoff from accepting a plea bargain of S_b is:

$$\pi_b^D = S_b + r^D\mu(G \mid b). \tag{TA.2}$$

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

$$\pi_R^D(t) = \rho^P\pi_T^D(t) + (1 - \rho^P)\pi_d^D, \tag{TA.3}$$

where $\pi_d^D = r^D\mu(G \mid d)$.

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

$$\pi_T^P = \nu(G \mid R)\{S_c(1 - F_G) - k^P - r_I^P\mu(I \mid c)(1 - F_G) - r_G^P\mu(G \mid a)F_G\}$$
$$+ \nu(I \mid R)\{S_c(1 - F_I) - k^P - r_I^P\mu(I \mid c)(1 - F_I) - r_G^P\mu(G \mid a)F_I\}. \tag{TA.4}$$

P's payoff from dropping the case is simply:

$$\pi_d^P = -r_G^P\mu(G \mid d). \tag{TA.5}$$

P's expected payoff following rejection is given by:

$$\pi_R^P = \rho^P\pi_T^P + (1 - \rho^P)\pi_d^P. \tag{TA.6}$$

A Preliminary Result

<u>Remark 1</u>. $\pi_T^D(I) < \pi_T^D(G)$. That is, an innocent defendant expects a smaller loss at trial than a guilty defendant (for given beliefs on the part of the observers and P).

Proof. First, note that for arbitrary positive values of ρ_G^D and ρ_I^D, the observers' posterior probability of guilt is higher following a conviction than following an acquittal if and only if $F_I > F_G$. More formally, $\mu(G \mid c)$ (>, =, <) $\mu(G \mid a)$ as F_I (>, =, <) F_G. To see this, notice that, by Bayes' Rule,

$$\mu(G \mid c) = \rho_G^D(1 - \lambda)(1 - F_G)/[\rho_G^D(1 - \lambda)(1 - F_G) + \rho_I^D\lambda(1 - F_I)], \quad \text{(TA.7)}$$

whereas

$$\mu(G \mid a) = \rho_G^D(1 - \lambda)F_G/[\rho_G^D(1 - \lambda)F_G + \rho_I^D\lambda F_I]. \quad \text{(TA.8)}$$

Simple though tedious algebra indicates that $\mu(G \mid c) > \mu(G \mid a)$ if and only if $F_I > F_G$, which is a maintained assumption. As ρ_I^D goes to zero, both $\mu(G \mid c)$ and $\mu(G \mid a)$ go to 1, whereas as ρ_G^D goes to zero, both $\mu(G \mid c)$ and $\mu(G \mid a)$ go to 0.

Second, the expression for $\pi_T^D(t)$ can be differentiated with respect to F_t to obtain:

$$\partial\pi_T^D(t)/\partial F_t = -S_c - r^D\{\mu(G \mid c) - \mu(G \mid a)\}.$$

Since the term in curly brackets has been proved to be non-negative, the entire expression is negative. Since $F_I > F_G$, the result follows. QED

Note that an innocent defendant and a guilty defendant expect the same loss if they accept the plea offer of S_b (for given beliefs on the part of the observers and P). This is because accepting the plea bargain results in case disposition b, which yields a payoff of $\pi_A^D = S_b + r^D\mu(G \mid b)$, independent of D's true type. S_b is the formal sanction, whereas $r^D\mu(G \mid b)$ is the informal sanction imposed by the observer, who believes that D's type is G with probability $\mu(G \mid b)$ if he accepts the plea bargain.

Maintained Restrictions

MR0. Given $(\rho_G^{D\Theta}, \rho_I^{D\Theta})$ and the corresponding beliefs for Θ, $S_c - r_I^P\mu(I \mid c) + r_G^P\mu(G \mid a) > 0$. This means that P strictly prefers to go to trial against a D she believes to be a G-type in comparison with one she believes to be an I-type.

To verify this implication of MR0, refer to equation (TA.4) and take the difference between the expressions in curly brackets. MR0 can also be interpreted to mean that P's payoff is increasing in $\nu(G \mid R)$, since $\nu(I \mid R) = 1 - \nu(G \mid R)$.

MR1. $(S_c - r^P_I)(1 - F_I) - k^P < 0$. This means that, if it were common knowledge (or commonly-believed) that D is innocent, then P would prefer to drop the case rather than proceed to trial.

To verify this implication of MR1, substitute $\nu(G \mid R) = \mu(G \mid a) = \mu(G \mid d) = 0$ and $\nu(I \mid R) = \mu(I \mid c) = 1$ into the formulas for π^P_T and π^P_d. This yields $\pi^P_T = S_c(1 - F_I) - k^P - r^P_I(1 - F_I)$, whereas $\pi^P_d = 0$. Then MR1 implies that $\pi^P_T < \pi^P_d$. To verify MR1(b), substitute $\nu(G \mid R) = 0$ into the formulas for π^P_T and π^P_d. This yields $\pi^P_T = S_c(1 - F_I) - k^P - r^P_I\mu(I \mid c)(1 - F_I) - r^P_G\mu(G \mid a)F_I$ and $\pi^P_d = -r^P_G\mu(G \mid d)$. Then MR1(b) implies that $\pi^P_T < \pi^P_d$.

MR2. $(1 - \lambda)[(S_c + r^P_G)(1 - F_G) - k^P] + \lambda[(S_c - r^P_I)(1 - F_I) - k^P] > 0$. This means that, if it were common knowledge (or commonly-believed) that the fraction of guilty defendants among those that rejected the plea offer is $1 - \lambda$, then P would prefer to take the case to trial rather than drop it.

To verify this implication of MR2, substitute $\nu(G \mid R) = 1 - \lambda$ and $\nu(I \mid R) = \lambda$ into the formula for π^P_T, equation (TA.4). Moreover, substitute $\mu(G \mid a) = (1 - \lambda)F_G/[(1 - \lambda)F_G + \lambda F_I]$ and $\mu(I \mid c) = \lambda(1 - F_I)/[(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]$ into that formula as well. This yields: $\pi^P_T = [(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]S^T - r^P_I\lambda(1 - F_I) - r^P_G(1 - \lambda)F_G - k^P$. Finally, substitute $\mu(G \mid d) = 1 - \lambda$ into the formula for π^P_d to obtain $\pi^P_d = -r^P_G(1 - \lambda)$. Then MR2 implies that $\pi^P_T > \pi^P_d$.

MR1 and MR2 imply the following, but we state it here for easy reference.

MR3. $(S_c + r^P_G)(1 - F_G) - k^P > 0$. This means that, if it were common knowledge (or commonly-believed) that D is guilty, then P would prefer to take the case to trial rather than drop it.

To verify this implication, substitute $\nu(G \mid R) = \mu(G \mid a) = \mu(G \mid d) = 1$ and $\nu(I \mid R) = \mu(I \mid c) = 0$ into the formulas for π_T^P and π_d^P. This yields $\pi_T^P = \{S_c(1 - F_G) - k^P - r_G^P F_G\}$, whereas $\pi_d^P = -r_G^P$. Then MR1 and MR2 imply that $\pi_T^P < \pi_d^P$.

The following remark considers arbitrary mixed strategies and the value to P of going to trial rather than dropping the case.

Remark 2. For arbitrary mixing probabilities (ρ_G^D, ρ_I^D), where ρ_t^D is the probability that type t rejects the plea offer, and the corresponding beliefs for both P and Θ, the expression $(\pi_T^P - \pi_d^P)$ (i.e., the difference between P's payoff from taking the case to trial and dropping it) is decreasing in ρ_I^D and increasing in ρ_G^D. Also, num$(\pi_T^P - \pi_d^P)$ is decreasing in ρ_I^D and increasing in ρ_G^D.

Proof. For arbitrary mixing probabilities (ρ_G^D, ρ_I^D), P's payoff from trial is given by:

$$\pi_T^P = \nu(G \mid R)\{S_c(1 - F_G) - k^P - r_I^P\mu(I \mid c)(1 - F_G) - r_G^P\mu(G \mid a)F_G\}$$
$$+ \nu(I \mid R)\{S_c(1 - F_I) - k^P - r_I^P\mu(I \mid c)(1 - F_I) - r_G^P\mu(G \mid a)F_I\}, \qquad \text{(TA.9)}$$

where $\nu(G \mid R) = \rho_G^D(1 - \lambda)/[\rho_G^D(1 - \lambda) + \rho_I^D\lambda]$; $\mu(I \mid c) = \rho_I^D\lambda(1 - F_I)/[\rho_G^D(1 - \lambda)(1 - F_G) + \rho_I^D\lambda(1 - F_I)]$; and $\mu(G \mid a) = \rho_G^D(1 - \lambda)F_G/[\rho_G^D(1 - \lambda)F_G + \rho_I^D\lambda F_I]$. P's payoff from dropping the case is given by $\pi_d^P = -r_G^P\mu(G \mid d)$, where $\mu(G \mid d) = \rho_G^D(1 - \lambda)/[\rho_G^D(1 - \lambda) + \rho_I^D\lambda]$. After much substitution, it can be shown that num$(\pi_T^P - \pi_d^P) = S_c[\rho_G^D(1 - \lambda)(1 - F_G) + \rho_I^D\lambda(1 - F_I)] - r_I^P\rho_I^D\lambda(1 - F_I) - k^P[\rho_G^D(1 - \lambda) + \rho_I^D\lambda] + r_G^P\rho_G^D(1 - \lambda)(1 - F_G)$ and denom$(\pi_T^P - \pi_d^P) = \rho_G^D(1 - \lambda) + \rho_I^D\lambda$. Differentiation, algebra, and MR1-MR3 yield the result that $(\pi_T^P - \pi_d^P)$ goes down as ρ_I^D goes up and $(\pi_T^P - \pi_d^P)$ goes up as ρ_G^D goes up. Considering only num$(\pi_T^P - \pi_d^P)$, this clearly goes down as ρ_I^D goes up (by MR1) and num$(\pi_T^P - \pi_d^P)$ goes up as ρ_G^D goes up (by MR3). QED

Candidates for Equilibria

There are 9 distinct candidate forms for equilibria. Candidates 1-4 are pure-strategy equilibria; that is, each type of D plays a particular strategy with probability 1. Candidates 5-9 involve at least one type of D mixing between accepting and rejecting the plea offer.

1. Types I and G accept the plea offer.
2. Type I rejects the plea offer, whereas type G accepts it.
3. Type I accepts the plea offer, whereas type G rejects it.
4. Types I and G reject the plea offer.
5. Type I rejects the plea offer, whereas type G mixes.
6. Type G accepts the plea offer, whereas type I mixes.
7. Type G rejects the plea offer, whereas type I mixes.
8. Type I accepts the plea offer, whereas type G mixes.
9. Both types mix.

We argue that the only candidate forms that can actually be equilibria are Candidates 4 and 5. We postpone characterization of these equilibria until after we dispose of those candidate forms that cannot be equilibria.

1. <u>Types I and G accept the plea offer</u>. In this putative equilibrium, the dispositions {a, c, d} are all out-of-equilibrium events. What should P believe if D unexpectedly rejects the plea offer? And what should observers believe if they unexpectedly observe a disposition of a, c or d? Since type I expects a lower loss from trial than type G, whereas both types expect the same loss from a dropped case, the equilibrium refinement D1 (Cho and Kreps, 1987) implies that unexpected rejection of the plea offer (and unexpected dispositions a, c or d) should be assigned to type I. This is because type I would be willing to risk a larger probability of trial ρ^P to defect from accepting to rejecting the plea offer than would type G. Formally, this means that in this putative equilibrium,

$\nu(G \mid R) = \mu(G \mid a) = \mu(G \mid c) = \mu(G \mid d) = 0$. Consequently, $\pi_T^P = S_c(1 - F_I) - k^P - r_I^P(1 - F_I)$ and $\pi_d^P = 0$; by MR1, this means that P will prefer to drop the case. Basically, if both types are expected to accept the plea bargain, then rejecting it is taken as a clear signal of innocence and P will therefore drop the case (P does not have to worry about informal sanctions from dropping the case, as observers take this disposition as a clear signal of innocence). But if P will drop the case following rejection of the plea offer, then both types will defect from this putative equilibrium to rejecting the plea offer. Thus, there cannot be an equilibrium of this form.

2. <u>Type I rejects the plea offer, whereas type G accepts it</u>. In this putative equilibrium, the observer's beliefs following the dispositions {a, c, d} (which could only occur following a rejection of the plea offer) are that D is surely of type I. Moreover, P also believes that a rejection implies type I. Thus, $\nu(G \mid R) = \mu(G \mid a) = \mu(G \mid c) = \mu(G \mid d) = 0$. Consequently, $\pi_T^P = S_c(1 - F_I) - k^P - r_I^P(1 - F_I)$ and $\pi_d^P = 0$; by MR1, this means that P will prefer to drop the case. Basically, if type G is expected to accept the plea bargain, then rejecting it is taken as a clear signal of innocence and P will therefore drop the case (P does not have to worry about informal sanctions from dropping the case, as observers take this disposition as a clear signal of innocence). Again, if P is expected to drop the case following a rejection of the plea offer, then the type G defendant will defect from this putative equilibrium to rejecting the plea offer. Thus, there cannot be an equilibrium of this form.

3. <u>Type I accepts the plea offer, whereas type G rejects it</u>. In this putative equilibrium, the observer's beliefs following the dispositions {a, c, d} (which could only occur following a rejection of the plea offer) are that D is surely of type G. Moreover, P also believes that a rejection implies type G. By MR3, P will prefer to take the case to trial rather than to drop it. Recall that – holding beliefs constant – type I faces a lower expected cost of trial than does type G, whereas they expect

the same loss by accepting the plea bargain. This implies that if type G prefers trial to the plea bargain (or is indifferent), then type I must strictly prefer trial to the plea bargain. Therefore, the type I defendant will defect from this putative equilibrium to rejecting the plea offer. Thus, there cannot be an equilibrium of this form.

6. Type G accepts the plea offer, whereas type I mixes. The argument is exactly the same as for candidate 2, and will be omitted.

7. Type G rejects the plea offer, whereas type I mixes. Recall that – holding beliefs constant – type I faces a lower expected cost of trial than does type G, whereas they expect the same loss by accepting the plea bargain (or by having the case dropped). If type I is indifferent between accepting the plea offer and rejecting it, then type G must strictly prefer the plea bargain, as long as P takes the case to trial with positive probability following rejection. Recall that MR2 ensures that if all type G's and all type I's are expected to reject the plea offer, then P would take the case to trial (rather than dropping it). Since the mixture of defendant types in this putative equilibrium puts more weight on type G (and less on type I) relative to the prior, Remark 2 implies that trial is even more attractive to P (relative to dropping the case), so P would take the case to trial. This implies that if type G prefers to reject the plea bargain (or is indifferent), then type I must strictly prefer to reject it. Therefore, the type I defendant will defect from this putative equilibrium to rejecting the plea offer. Thus, there cannot be an equilibrium of this form.

8. Type I accepts the plea offer, whereas type G mixes. In this putative equilibrium, any rejection of the plea offer, and any outcome {a, c, d} are attributed to type G. By MR3, P would take the case to trial following a rejection. If type G is mixing, then he must be indifferent between the plea offer and trial. But since type I expects a smaller loss than type G from trial (but the same loss as G from

accepting the plea bargain), it must be that type I strictly prefers trial to accepting the plea bargain. Therefore, the type I defendant will defect from this putative equilibrium to rejecting the plea offer. Thus, there cannot be an equilibrium of this form.

9. Both types mix. Again – holding beliefs constant – type I faces a lower expected cost of trial than does type G, whereas they expect the same loss by accepting the plea bargain (or by having the case dropped). If type I is indifferent between accepting the plea offer and rejecting it, then type G must strictly prefer the plea bargain, as long as P takes the case to trial with positive probability following rejection, which would cause type G to defect from this putative equilibrium to accepting the plea.

Only if P drops the case with probability one following rejection can both defendant types be made indifferent. Such an equilibrium can be ruled out if we assume that, when both types are indifferent, they reject the plea offer (or are believed, by both P and the observers, to reject the plea offer) with the same probability. In this case, the mixture of rejecting types is (believed to be) the same as the prior, and P prefers to take the case to trial rather than dropping it. But then it cannot be that both D types are indifferent, as type I expects a smaller loss from trial than type G. Alternatively, a stronger version of MR2 could guarantee that P prefers to make a demand that is unacceptable to both types, rather than dropping the case against both types. While provoking rejection by both types does not change the observers' beliefs (because rejection is on the equilibrium path), it does change P's posterior beliefs. Her payoff from trial is now:

$$\tilde{\pi}_T^P = (1 - \lambda)\{S_c(1 - F_G) - k^P - r_I^P\mu(I \mid c)(1 - F_G) - r_G^P\mu(G \mid a)F_G\}$$
$$+ \lambda\{S_c(1 - F_I) - k^P - r_I^P\mu(I \mid c)(1 - F_I) - r_G^P\mu(G \mid a)F_I\},$$

where $\mu(I \mid c) = \rho_I^D\lambda(1 - F_I)/[\rho_G^D(1 - \lambda)(1 - F_G) + \rho_I^D\lambda(1 - F_I)]$ and $\mu(G \mid a) = \rho_G^D(1 - \lambda)F_G/[\rho_G^D(1 - \lambda)F_G$

$+ \rho_I^D\lambda F_I]$. P's payoff from dropping the case is still given by $\pi_d^P = -r_G^P\mu(G \mid d)$, where $\mu(G \mid d) = \rho_G^D(1 - \lambda)/[\rho_G^D(1 - \lambda) + \rho_I^D\lambda]$. If $[(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]S_c$ is assumed to be sufficiently large, then P will prefer to provoke trial against both types rather than dropping the case against both types. This will also undermine a putative equilibrium wherein both types of defendant mix between accepting and rejecting the plea offer.

Characterizing Equilibrium

The only remaining candidate forms for an equilibrium are Candidates 4 (types I and G reject the plea offer) and 5 (type I rejects the plea offer, whereas type G mixes). Thus, an innocent defendant rejects the plea offer with probability one, but a guilty defendant may accept the plea offer with positive probability. Because Candidate 4 is a limiting case of Candidate 5, we can focus on Candidate 5.

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_b, 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; type I will always reject the plea offer in this putative equilibrium). 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. 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 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 $\pi_R^D(G; \rho_G^{D\Theta}) = \rho^P\pi_T^D(G; \rho_G^{D\Theta}) + (1 - \rho^P)\pi_d^D(\rho_G^{D\Theta}) = \pi_b^D(\rho_G^{D\Theta})$. That is, if: $\rho^P[\pi_T^D(G; \rho_G^{D\Theta}) - \pi_d^D(\rho_G^{D\Theta})] = \pi_b^D(\rho_G^{D\Theta}) - \pi_d^D(\rho_G^{D\Theta})$. Substitution and simplification yields:

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

Upon collecting terms, the value of ρ^P that results in this equality is:

$$\rho^P(S_b; \rho_G^{D\Theta}) = \frac{\{S_b + r^D(1 - \mu(G \mid d; \rho_G^{D\Theta}))\}}{\{S_c(1 - F_G) + k^D + r^D[\mu(G \mid c; \rho_G^{D\Theta})(1 - F_G) + \mu(G \mid a; \rho_G^{D\Theta})F_G - \mu(G \mid d; \rho_G^{D\Theta})]\}}. \tag{TA.10}$$

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. 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 below), which implies that type G would prefer that P drop the case against him rather than take it to trial.

Remark 3. The denominator of the expression $\rho^P(S_{b;}\, \rho_G^{D\Theta})$ is positive.

Proof. A sufficient condition for the denominator to be positive is that

$$\mu(G \mid c; \rho_G^{D\Theta})(1 - F_G) + \mu(G \mid a; \rho_G^{D\Theta})F_G - \mu(G \mid d; \rho_G^{D\Theta})$$
$$= [\mu(G \mid c; \rho_G^{D\Theta}) - \mu(G \mid d; \rho_G^{D\Theta})](1 - F_G) + [\mu(G \mid a; \rho_G^{D\Theta}) - \mu(G \mid d; \rho_G^{D\Theta})]F_G > 0.$$

Recall that $\mu(G \mid c; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)(1 - F_G)/[\rho_G^{D\Theta}(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]$; $\mu(G \mid a; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)F_G/[\rho_G^{D\Theta}(1 - \lambda)F_G + \lambda F_I]$; and $\mu(G \mid d; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \lambda]$. Let $X = [\rho_G^{D\Theta}(1 - \lambda)(1 - F_G)$

$+ \lambda(1 - F_I)]$ and let $Y = [\rho_G^{D\Theta}(1 - \lambda)F_G + \lambda F_I]$; so $X + Y = [\rho_G^{D\Theta}(1 - \lambda) + \lambda]$. Then:

$$[\mu(G \mid c; \rho_G^{D\Theta}) - \mu(G \mid d; \rho_G^{D\Theta})](1 - F_G) + [\mu(G \mid a; \rho_G^{D\Theta}) - \mu(G \mid d; \rho_G^{D\Theta})]F_G$$

$$= \{[\rho_G^{D\Theta}(1 - \lambda)(1 - F_G)/X] - [\rho_G^{D\Theta}(1 - \lambda)/(X + Y)]\}(1 - F_G)$$

$$+ \{[\rho_G^{D\Theta}(1 - \lambda)F_G/Y] - [\rho_G^{D\Theta}(1 - \lambda)/(X + Y)]\}F_G > 0$$

if and only if (after some algebra) $[(1 - F_G)(X + Y)](1 - F_G)Y + [F_G(X + Y) - Y]F_G X > 0$, which can be verified by substituting for X, Y and X + Y, collecting terms, and recalling that $F_I > F_G$. QED

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 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:

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

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

$$= - r_G^P \mu(G \mid d; \rho_G^{D\Theta}). \quad \text{(TA.11)}$$

Substituting for the beliefs and simplifying yields (see also the proof of Remark 2):

$$\text{num}(\pi_T^P(\rho_G^{D\Theta}) - \pi_d^P(\rho_G^{D\Theta})) = S_c[\rho_G^{D\Theta}(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)] - r_I^P\lambda(1 - F_I)$$

$$- k^P[\rho_G^{D\Theta}(1 - \lambda) + \lambda] + r_G^P\rho_G^{D\Theta}(1 - \lambda)(1 - F_G), \quad \text{(TA.12)}$$

and $\text{denom}(\pi_T^P(\rho_G^{D\Theta}) - \pi_d^P(\rho_G^{D\Theta})) = [\rho_G^{D\Theta}(1 - \lambda) + \lambda]$. The expression $\text{num}(\pi_T^P(\rho_G^{D\Theta}) - \pi_d^P(\rho_G^{D\Theta}))$ is increasing in $\rho_G^{D\Theta}$ by MR3. Moreover, we know from MR1 that this expression is negative for $\rho_G^{D\Theta} = 0$ (that is, when a D of type G is never expected to reject), and we know from MR2 that it is positive for $\rho_G^{D\Theta} = 1$ (that is, when a D of type G is always expected to reject). Therefore, the unique value of $\rho_G^{D\Theta}$ that will make P indifferent between trying and dropping the case is given by:

$$\rho_G^{D0} = - \lambda[(S_c - r_I^P)(1 - F_I) - k^P]/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P], \quad \text{(TA.13)}$$

where the numerator is positive by MR1; the denominator is positive by MR3; and the ratio is a fraction by MR2. 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) believes that type G rejects the plea offer with probability ρ_G^{D0}. Thus, the mixed-strategy 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:

$$(1 - \rho_G^{D0})(1 - \lambda)S_b + (\rho_G^{D0}(1 - \lambda) + \lambda)[\rho^P(S_{b;}\, \rho_G^{D0})\pi_T^P(\rho_G^{D0}) + (1 - \rho^P(S_{b;}\, \rho_G^{D0}))\pi_d^P(\rho_G^{D0})]. \quad \text{(TA.14)}$$

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, $r^D[1 - \mu(G \mid c; \rho_G^{D0})(1 - F_G) - \mu(G \mid a; \rho_G^{D0})F_G] \leq S_c(1 - F_G) + k^D$. Since the term in brackets on the left-hand-side can be re-written as $(1 - F_G)(1 - \mu(G \mid c; \rho_G^{D0})) + F_G(1 - \mu(G \mid a; \rho_G^{D0}))$, it is clearly positive.

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

The expression $r^D[1 - \mu(G \mid c; \rho_G^{D0})(1 - F_G) - \mu(G \mid a; \rho_G^{D0})F_G]$ is the increment in informal sanctions that the D of type G suffers by accepting a plea (which only a true G is expected to do) rather than going to trial (where there is a chance of conviction and a chance of acquittal, with corresponding informal sanctions). If there were no informal sanctions for D, then Condition 1 would be satisfied automatically. Thus, informal sanctions on D constrain P's ability to settle cases via plea bargain.

Returning to P's payoff as a function of S_b (i.e., equation (TA.14)), 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 the upper limit of the feasible range, $S_b(\rho_G^{D0}) = \pi_T^D(G; \rho_G^{D0}) - r^D$; this is rejected by type G with probability

ρ_G^{D0}, and P (though indifferent) 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 G-types accept, and others reject, the offer, with the optimal offer within this set being at the upper endpoint. But P could make a higher demand that would provoke certain rejection. 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 the levels implied by ρ_G^{D0}, because trial is already on the equilibrium path). Thus, P prefers (at least weakly) the hypothesized equilibrium to defection as long as:

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

$$\geq S_c(1 - F_G) - k^P - r_I^P\mu(I \mid c; \rho_G^{D0})(1 - F_G) - r_G^P\mu(G \mid a; \rho_G^{D0})F_G. \qquad \text{(TA.15)}$$

Rearranging, we can write this as:

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

Condition 2. For P to find it preferable to settle with a D of type G rather than provoking a trial, it must be that:

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

Again, if D faced no informal sanctions, then Condition 2 would be satisfied. High informal sanctions on D can undermine P's desire to settle using plea bargaining.

Finally, P could also defect by dropping all cases (resulting in Θ applying the out-of-equilibrium beliefs $\mu(G \mid d;\, \rho_G^{D0})$); 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:

$$(1 - \rho_G^{D0})(1 - \lambda)[\pi_T^D(G;\, \rho_G^{D0}) - r^D] + (\rho_G^{D0}(1 - \lambda) + \lambda)\pi_T^P(\rho_G^{D0}). \qquad \text{(TA.16)}$$

We already know that $\pi_T^P(\rho_G^{D0}) = \pi_d^P(\rho_G^{D0})$ by construction (and $\pi_T^P(\rho_G^D) > \pi_d^P(\rho_G^D)$ for $\rho_G^D > \rho_G^{D0}$). 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. Both Conditions 1 and 2 are restrictions on r^D; however, we have been unable to determine which right-hand-side provides the tighter constraint.

Later in this Technical Appendix, we will argue that the equilibrium we have just characterized is the unique equilibrium. But for now we continue to investigate the comparative statics of this equilibrium.

Comparative Statics

Here we summarize comparative static effects of parameter changes in r_I^P, r_G^P, r^D, k^P, k^D, F_G, F_I, λ, and S_c on equilibrium strategies such as the plea offer and the likelihood of plea bargaining success. Recall that the likelihood of plea bargaining failure is ρ_G^{D0}, where

$$\rho_G^{D0} = - \lambda[(S_c - r_I^P)(1 - F_I) - k^P]/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P], \qquad \text{(TA.17)}$$

and the equilibrium plea offer is

$$S_b(\rho_G^{D0}) = \pi_T^D(G;\, \rho_G^{D0}) - r^D = S_c(1 - F_G) + k^D + r^D\mu(G \mid c;\, \rho_G^{D0})(1 - F_G) + r^D\mu(G \mid a;\, \rho_G^{D0})F_G - r^D. \text{(TA.18)}$$

First, we consider the impact of changes in the parameters on ρ_G^{D0}. Recall that a higher value of ρ_G^D makes trial more attractive to P (relative to dropping the case) following a rejection, and ρ_G^{D0} makes P indifferent between these two decisions (even though she goes to trial with probability one in equilibrium). Therefore, any parameter change that would tip P toward one decision or the other must be counter-balanced by a change in ρ_G^{D0} that restore's P's indifference.

$$\partial\rho_G^{D0}/\partial\lambda = -[(S_c - r_I^P)(1 - F_I) - k^P]/(1 - \lambda)^2[(S_c + r_G^P)(1 - F_G) - k^P] > 0;$$

$$\partial\rho_G^{D0}/\partial r_I^P = \lambda(1 - F_I)/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P] > 0;$$

$$\partial\rho_G^{D0}/\partial k^P = -\lambda\{[(S_c - r_I^P)(1 - F_I) - k^P] - [(S_c + r_G^P)(1 - F_G) - k^P]\}/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P]^2 > 0;$$

$$\partial\rho_G^{D0}/\partial F_G = -\lambda[(S_c - r_I^P)(1 - F_I) - k^P](S_c + r_G^P)/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P]^2 > 0.$$

An increase in λ (the fraction of innocent among those arrested), r_I^P (the sanction rate for punishing an innocent defendant), k^P (P's cost of trial), or F_G (the probability that a guilty defendant is acquitted) has the direct effect of making trial less attractive, so the fraction of guilty types in the pool of those rejecting must increase to restore P's willingness to go to trial.

$$\partial\rho_G^{D0}/\partial r_G^P = \lambda[(S_c - r_I^P)(1 - F_I) - k^P](1 - F_G)/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P]^2 < 0.$$

$$\partial\rho_G^{D0}/\partial S_c = \frac{\{(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P](-\lambda(1 - F_I)) + \lambda[(S_c - r_I^P)(1 - F_I) - k^P](1 - \lambda)(1 - F_G)\}}{\{(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P]\}^2} < 0.$$

An increase in r_G^P (the sanction rate for failing to punish a guilty defendant) or S_c (the formal sanction) has the direct effect of making trial more attractive (relative to dropping the case), so the fraction of guilty types in the pool of those rejecting can decrease and yet maintain P's willingness to go to trial. Some comparative statics are ambiguous (i.e., they could go either way, depending on the relative magnitude of parameters) or we are unable to determine the direction of the impact.

$$\partial\rho_G^{D0}/\partial F_I = \lambda(S_c - r_I^P)/(1 - \lambda)[(S_c + r_G^P)(1 - F_G) - k^P] > 0 \text{ (resp., } < 0\text{) if } (S_c - r_I^P) > 0 \text{ (resp., } < 0\text{)};$$

Finally, as r^D and k^D do not appear in P's payoffs, they do not have an impact on ρ_G^{D0}; that is, $\partial\rho_G^{D0}/\partial k^D = 0$ and $\partial\rho_G^{D0}/\partial r^D = 0$.

Now consider the impact of parameter changes on $S_b(\rho_G^{D0})$, which is an increasing function. There are several parameters that affect the equilibrium plea offer only indirectly through ρ_G^{D0}.

$\partial S_b(\rho_G^{D0})/\partial\lambda = S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial\lambda) > 0;$

$\partial S_b(\rho_G^{D0})/\partial r_I^P = S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial r_I^P) > 0;$

$\partial S_b(\rho_G^{D0})/\partial k^P = S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial k^P) > 0;$

$\partial S_b(\rho_G^{D0})/\partial r_G^P = S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial r_G^P) < 0;$

$\partial S_b(\rho_G^{D0})/\partial F_I = S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial F_I) > 0$ (resp., < 0) if $(S_c - r_I^P) > 0$ (resp., < 0).

The parameters r^D and k^D affect the equilibrium offer only directly, as ρ_G^{D0} does not depend on them.

$\partial S_b(\rho_G^{D0})/\partial r^D = \mu(G \mid c; \rho_G^{D0})(1 - F_G) + \mu(G \mid a; \rho_G^{D0})F_G - 1 < 0;$

$\partial S_b(\rho_G^{D0})/\partial k^D = 1 > 0;$

The parameters F_G and S_c affect the plea offer both directly and indirectly.

$\partial S_b(\rho_G^{D0})/\partial S_c = (1 - F_G) + S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial S_c) = ???$, as the direct effect is positive and the indirect effect is negative.

$\partial S_b(\rho_G^{D0})/\partial F_G = \{- S_c - r^D[\mu(G \mid c; \rho_G^{D0}) - \mu(G \mid a; \rho_G^{D0})]\} + S_b'(\rho_G^{D0})(\partial\rho_G^{D0}/\partial F_G) = ???$, as the direct effect (in curly brackets) is negative and the indirect effect is positive.

Recall that the right-hand-side of Condition 1 is:

$[S_c(1 - F_G) + k^D]/[1 - \mu(G \mid c; \rho_G^{D0})(1 - F_G) - \mu(G \mid a; \rho_G^{D0})F_G].$

Since the denominator is decreasing in ρ_G^{D0}, and ρ_G^{D0} is increasing (resp., decreasing) in r_I^P (resp., r_G^P), it follows that the r.h.s. of Condition 1 is increasing in r_I^P and decreasing in r_G^P.

The right-hand-side of Condition 2 is:

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

Again, the denominator is decreasing in r_I^P, and decreasing in r_G^P. Since the numerator is increasing in r_I^P (the only questionable term is $r_I^P\mu(I \mid c; \rho_G^{D0})$ because $\mu(I \mid c; \rho_G^{D0})$ is decreasing, but overall this product is increasing in r_I^P), it follows that the r.h.s. of Condition 2 is increasing in r_I^P. We are unable to determine the impact of an increase in r_G^P on the r.h.s. of Condition 2, because both the numerator and the denominator are increasing in r_G^P.

Finally, we show that an increase in S_c decreases the expected informal sanctions facing P and a D of type I, and increases the expected informal sanctions facing a D of type G. To see this, we first consider the ex ante expected informal sanctions facing P. This is given by:

$$(1 - \lambda)\rho_G^{D0}\{r_I^P\mu(I \mid c; \rho_G^{D0})(1 - F_G) + r_G^P\mu(G \mid a; \rho_G^{D0})F_G\} + \lambda\{r_I^P\mu(I \mid c; \rho_G^{D0})(1 - F_I) + r_G^P\mu(G \mid a; \rho_G^{D0})F_I\}.$$

Upon substituting for the beliefs and collecting terms, this reduces to $r_I^P\lambda(1 - F_I) + r_G^P(1 - \lambda)\rho_G^{D0}F_G$. Since this expression is increasing in ρ_G^{D0}, which is itself decreasing in S_c, it follows that P's ex ante expected informal sanctions are decreasing in S_c. Next, we consider the ex ante expected informal sanctions facing a D of unknown type. This is given by:

$$r^D\lambda\{(1 - F_I)\mu(G \mid c; \rho_G^{D0}) + F_I\mu(G \mid a; \rho_G^{D0})\}$$
$$+ r^D(1 - \lambda)\{(1 - \rho_G^{D0}) + \rho_G^{D0}(1 - F_G)\mu(G \mid c; \rho_G^{D0}) + \rho_G^{D0}F_G\mu(G \mid a; \rho_G^{D0})\}.$$

The terms in the first line represent the contribution to ex ante expected informal sanctions generated by the D of type I (who never accepts the plea), and the terms in the second line represent the contribution generated by the D of type G (who sometimes accepts the plea). Collecting coefficients on the expressions $\mu(G \mid c; \rho_G^{D0})$ and $\mu(G \mid a; \rho_G^{D0})$ yields the following version:

$$r^D\{\lambda(1 - F_I) + (1 - \lambda)\rho_G^{D0}(1 - F_G)\}\mu(G \mid c; \rho_G^{D0}) + r^D\{\lambda F_I + (1 - \lambda)\rho_G^{D0}F_G\}\mu(G \mid a; \rho_G^{D0})\} + r^D(1 - \lambda)(1 - \rho_G^{D0}).$$

Upon recalling the form of the beliefs, $\mu(G \mid c; \rho_G^{D0})$ and $\mu(G \mid a; \rho_G^{D0})$, we see that the expressions in curly brackets above are the same as the denominators in the beliefs they multiply; thus, the expression above reduces to:

$$r^D\{(1 - \lambda)\rho_G^{D0}(1 - F_G)\} + r^D\{(1 - \lambda)\rho_G^{D0}F_G\} + r^D(1 - \lambda)(1 - \rho_G^{D0}) = r^D(1 - \lambda).$$

That is, the ex ante expected informal sanctions facing a D of unknown type are <u>independent</u> of S_c.

Since it is clear that the expected informal sanctions facing a D of type I, which are given by $r^D\{(1 - F_I)\mu(G \mid c; \rho_G^{D0}) + F_I\mu(G \mid a; \rho_G^{D0})\}$, are decreasing in S_c (as $\mu(G \mid c; \rho_G^{D0})$ and $\mu(G \mid a; \rho_G^{D0})$ are both increasing in ρ_G^{D0} and ρ_G^{D0} is decreasing in S_c), then the fact that the ex ante expected informal sanctions facing a D of unknown type are independent of S_c implies that the expected informal sanctions facing a D of type G, given by $r^D\{(1 - \rho_G^{D0}) + \rho_G^{D0}(1 - F_G)\mu(G \mid c; \rho_G^{D0}) + \rho_G^{D0}F_G\mu(G \mid a; \rho_G^{D0})\}$, must be increasing in S_c.

Uniqueness of Equilibrium

Recall equation (TA.12), which represented $num(\pi_T^P(\rho_G^{D\Theta}) - \pi_d^P(\rho_G^{D\Theta}))$ for a putative equilibrium wherein both Θ and P held common conjectures $\rho_G^{D\Theta}$ about the probability that a G-type would reject the plea offer. This expression is increasing in $\rho_G^{D\Theta}$ by MR3. Moreover, we know from MR1 that this expression is negative for $\rho_G^{D\Theta} = 0$ (that is, when a D of type G is never expected to reject), and we know from MR2 that it is positive for $\rho_G^{D\Theta} = 1$ (that is, when a D of type G is always expected to reject). Finally, we know that it is just equal to zero at ρ_G^{D0}. Thus, MR1 and MR2 imply that there cannot be an equilibrium wherein $\rho_G^D < \rho_G^{D0}$ (for, if P and Θ held these conjectures, then P would prefer to drop the case following a rejection rather than taking it to trial).

However, it appears that there <u>could</u> be an equilibrium at some $\hat{\rho}_G^D > \rho_G^{D0}$. If there were such an equilibrium, it would involve both P and Θ holding the conjecture $\hat{\rho}_G^D$ and P making the plea offer of $S_b(\hat{\rho}_G^D) = \pi_T^D(G; \hat{\rho}_G^D) - r^D$. Given this plea offer (and anticipating Θ's conjecture), type G would be willing to randomize using the rejection probability $\hat{\rho}_G^D$, and P would strictly prefer to take the case to trial following a rejection. We now argue that P has a profitable deviation which involves offering a slightly lower plea offer and obtaining a greater frequency of plea acceptance.

First, recall MR0, which says that for given beliefs on the part of Θ, P's payoff is increasing in her beliefs $\nu(G \mid R)$ that a rejecting D is a G-type. For $\rho_G^{D\Theta} = \rho_G^{D0}$, since P is just indifferent about trial versus dropping the case when she also believes that $\nu(G \mid R) = \rho_G^{D0}$, it must be that P's payoff against a G-type strictly warrants going to trial whereas P's payoff against an I-type strictly warrants dropping the case; it is the mixture of types that makes her indifferent.

Next consider $\rho_G^{D\Theta} = \hat{\rho}_G^{D} > \rho_G^{D0}$. Now P strictly prefers trial to dropping the case when she also believes that $\nu(G \mid R) = \hat{\rho}_G^{D}$. But since she is not indifferent, we cannot rule out the possibility that P would prefer trial to dropping the case even if she believed that $\nu(G \mid R) = 0$.

Suppose that $\rho_G^{D\Theta} = \hat{\rho}_G^{D} > \rho_G^{D0}$, and $\nu(G \mid R) = 0$. Then P's net expected payoff from proceeding to trial rather than dropping the case is:

$$S_c(1 - F_I) - k^P - r_I^P\mu(I \mid c)(1 - F_I) - r_G^P\mu(G \mid a)F_I + r_G^P\mu(G \mid d), \tag{TA.19}$$

where Θ's beliefs are constructed using the conjecture $\hat{\rho}_G^{D}$. If this expression is non-negative, then P could deviate to a plea offer of $S_b(\hat{\rho}_G^{D}) - \varepsilon$ and G – knowing that P would continue to have a credible threat of trial even if she subsequently believed that any rejecting D is surely an I-type – will accept this plea offer for sure (as he is no longer indifferent between accepting and rejecting, knowing that a rejection will still result in trial). Thus, a vanishingly small cut in the plea offer would result in a discrete increase in the plea acceptance rate. This provides a profitable deviation for P.

On the other hand, suppose that the expression in (TA.19) is negative. Then if P were to cut the plea offer slightly, the G-type could not be confident that P would necessarily proceed to trial following a rejection of this slightly-reduced offer. Indeed, the G-type would know that, in order for P to have a credible threat of trial following a rejected offer, P must keep a sufficiently high fraction of G-types among those rejecting the plea offer. We now argue that, <u>holding Θ's beliefs constant</u> (that is, as constructed using the conjecture $\hat{\rho}^D_G$), the continuation equilibrium following a plea offer of $S_b(\hat{\rho}^D_G)$ - ε is a mixed-strategy equilibrium between P and the G-type wherein P mixes between trial and dropping the case and the G-type mixes between accepting and rejecting the plea offer. In particular, if type G were to accept the slightly lower offer for sure, then P would certainly drop the case following a rejection; but then the G-type would prefer not to accept the slightly lower offer, as he would expect his case to be dropped. So there is no equilibrium in which the G-type plays a pure strategy.

In order to make the G-type indifferent, P must employ a probability of trial of ρ^P such that:

$$S_b(\hat{\rho}^D_G) - \varepsilon + r^D = \rho^P \pi^D_T(G;\, \hat{\rho}^D_G) + (1 - \rho^P) r^D \mu(G \mid d);$$

that is, $\rho^P(S_b(\hat{\rho}^D_G) - \varepsilon) = [S_b(\hat{\rho}^D_G) - \varepsilon + r^D - r^D\mu(G \mid d)]/[\pi^D_T(G;\, \hat{\rho}^D_G) - r^D\mu(G \mid d)]$. Note that as ε goes to zero, $\rho^P(S_b(\hat{\rho}^D_G) - \varepsilon)$ goes to 1. Having cut her plea offer slightly, P has to offer G a small probability of having his case dropped if he is to be made indifferent between accepting and rejecting the plea offer.

On the other hand, in order to make P indifferent between trial and dropping the case, the G-type must reject the plea offer with probability $\bar{\rho}^D_G$ such that (employing the resulting beliefs $\nu(G \mid R) = \bar{\rho}^D_G(1 - \lambda)/[\bar{\rho}^D_G(1 - \lambda) + \lambda]$ in equation (TA.9)) the expression below equals $-r^P_G\mu(G \mid d)$.

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

Recall that the expression above strictly exceeds $-r^P_G\mu(G \mid d)$ when P has the same conjectures as Θ, so that P believes $\nu(G \mid R) = \hat{\rho}^D_G(1 - \lambda)/[\hat{\rho}^D_G(1 - \lambda) + \lambda]$. By MR0, the way to move P from strictly preferring trial to being indifferent is to lower the fraction of G-types among those rejecting the plea offer. Thus, it follows that $\bar{\rho}^{D}_{G} < \hat{\rho}^{D}_{G}$. Since $\bar{\rho}^{D}_{G}$ is independent of ε, we have found that P can make a vanishingly small cut in the plea offer, and drop a vanishingly small fraction of cases following rejection (both costs to P) but will thereby obtain a discrete increase in the plea acceptance rate. This represents a profitable deviation for P from any putative equilibrium involving $\hat{\rho}^{D}_{G} > \rho^{D0}_{G}$.

Equilibrium Plea Acceptance by Innocent Defendants

In the base model, D has two possible types: G and I. When we add the idea of a strong (S) and a weak (W) version of D, with ω being the probability that D is weak, we end up with four possible types: GS, GW, IS, and IW. For the base model, equations (A1)-(A4) in the Appendix describe Θ's posterior belief that D is G, given the case disposition a, b, c, or d. These depend on Θ's conjectures about the probability that a D of type G (resp., I) would reject the plea offer, which is denoted by $\rho^{D\Theta}_{G}$ (resp., $\rho^{D\Theta}_{I}$). When we expand our type set as above, we will need four type-indexed probabilities of rejection: $\rho^{D\Theta}_{GS}$ denotes Θ's conjectures about the probability that a D of type GS would reject the plea offer. The expressions $\rho^{D\Theta}_{GW}$, $\rho^{D\Theta}_{IS}$, and $\rho^{D\Theta}_{IW}$ are similarly defined. The relevant equations, for arbitrary conjectures, are modified as follows:

$\mu(G \mid a) = [\omega\rho^{D\Theta}_{GW}+(1-\omega)\rho^{D\Theta}_{GS}](1 - \lambda)F_G/\{[\omega\rho^{D\Theta}_{GW}+(1-\omega)\rho^{D\Theta}_{GS}](1 - \lambda)F_G + [\omega\rho^{D\Theta}_{IW}+(1-\omega)\rho^{D\Theta}_{IS}]\lambda F_I\}$;

$\mu(G \mid b) = [\omega(1-\rho^{D\Theta}_{GW})+(1-\omega)(1-\rho^{D\Theta}_{GS})](1-\lambda)/\{[\omega(1-\rho^{D\Theta}_{GW})+(1-\omega)(1-\rho^{D\Theta}_{GS})](1-\lambda)+[\omega(1-\rho^{D\Theta}_{IW})+(1-\omega)(1-\rho^{D\Theta}_{IS})]\lambda\}$;

$\mu(G \mid c) = [\omega\rho^{D\Theta}_{GW}+(1-\omega)\rho^{D\Theta}_{GS}](1-\lambda)(1-F_G)/\{[\omega\rho^{D\Theta}_{GW}+(1-\omega)\rho^{D\Theta}_{GS}](1-\lambda)(1-F_G) + [\omega\rho^{D\Theta}_{IW}+(1-\omega)\rho^{D\Theta}_{IS}]\lambda(1-F_I)\}$;

and $\mu(G \mid d) = [\omega\rho^{D\Theta}_{GW}+(1-\omega)\rho^{D\Theta}_{GS}](1 - \lambda)/\{[\omega\rho^{D\Theta}_{GW}+(1-\omega)\rho^{D\Theta}_{GS}](1 - \lambda) + [\omega\rho^{D\Theta}_{IW}+(1-\omega)\rho^{D\Theta}_{IS}]\lambda\}$.

As long as the share of weak types (ω) is sufficiently small, the equilibrium will still involve P making a plea offer that renders the D of type GS indifferent between acceptance and going to trial; the D of type I rejects this demand for sure ($\rho^{D\Theta}_{IS} = 1$) and both weak types accept it for sure ($\rho^{D\Theta}_{GW} =$

$\rho_{IW}^{D\Theta} = 0$). Substituting these into the equations above gives the following:

$$\mu(G \mid a) = \rho_{GS}^{D\Theta}(1 - \lambda)F_G/[\rho_{GS}^{D\Theta}(1 - \lambda)F_G + \lambda F_I]; \tag{TA.20a}$$

$$\mu(G \mid b) = [\omega + (1 - \omega)(1 - \rho_{GS}^{D\Theta})](1 - \lambda)/\{[\omega + (1 - \omega)(1 - \rho_{GS}^{D\Theta})](1 - \lambda) + \omega\lambda\}; \tag{TA.20b}$$

$$\mu(G \mid c) = \rho_{GS}^{D\Theta}(1 - \lambda)(1 - F_G)/[\rho_{GS}^{D\Theta}(1 - \lambda)(1 - F_G) + \lambda(1 - F_I)]; \tag{TA.20c}$$

and $$\mu(G \mid d) = \rho_{GS}^{D\Theta}(1 - \lambda)/[\rho_{GS}^{D\Theta}(1 - \lambda) + \lambda]. \tag{TA.20d}$$

It is straightforward to compare these equations with equations (A1)-(A4), after substituting therein the equilibrium value $\rho_I^{D\Theta} = 1$. In particular, the system of equations above is the same as the system (A1)-(A4), with one exception. In the base model, equation (A2) becomes $\mu(G \mid b) = 1$; the acceptance of a plea offer is a clear signal of guilt. Whereas equation (TA.20b) provides a value for $\mu(G \mid b)$ that is less than 1 for all $\omega > 0$; that is, acceptance of a plea offer is no longer a sure sign of guilt, as a fraction ω of innocent defendants also accept the plea offer. Moreover, $\mu(G \mid b)$ is a decreasing function of ω; as the fraction of weak defendants increases, accepting the plea offer is an increasingly weak signal of guilt.

Now consider P's beliefs upon observing a rejection of her plea offer. In the base model, for arbitrary conjectures these beliefs are given by: $\nu(G \mid R; \rho_G^{D\Theta}, \rho_I^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \rho_I^{D\Theta}\lambda]$ (recall, P and Θ have common conjectures about D). In equilibrium, $\rho_I^{D\Theta} = 1$, so $\nu(G \mid R; \rho_G^{D\Theta}, \rho_I^{D\Theta} = 1) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \lambda]$. With four D types, this expression becomes:

$$\nu(G \mid R; \rho_{GS}^{D\Theta}, \rho_{GW}^{D\Theta}, \rho_{IS}^{D\Theta}, \rho_{IW}^{D\Theta})$$
$$= [\omega\rho_{GW}^{D\Theta} + (1 - \omega)\rho_{GS}^{D\Theta}](1 - \lambda)/\{[\omega\rho_{GW}^{D\Theta} + (1 - \omega)\rho_{GS}^{D\Theta}](1 - \lambda) + [\omega\rho_{IW}^{D\Theta} + (1 - \omega)\rho_{IS}^{D\Theta}]\lambda\}.$$

In equilibrium, $\rho_{IS}^{D\Theta} = 1$ and $\rho_{GW}^{D\Theta} = \rho_{IW}^{D\Theta} = 0$; making these substitutions results in: $\nu(G \mid R; \rho_{GS}^{D\Theta}, \rho_{GW}^{D\Theta} = 0, \rho_{IS}^{D\Theta} = 1, \rho_{IW}^{D\Theta} = 0) = \rho_{GS}^{D\Theta}(1 - \lambda)/[\rho_{GS}^{D\Theta}(1 - \lambda) + \lambda]$. This is exactly the same form as in the base model. Given that the fraction ω of both G-types and I-types accept the plea offer for sure, and all of the remaining (i.e., strong) I-types reject the plea offer, the mixture of innocent and guilty defendants among those that rejected the plea offer has exactly the same form.

This allows us to write P's indifference condition between taking a case to trial versus dropping it following rejection as follows (this is simply equation (TA.11) with $\rho_{GS}^{D\Theta}$ in place of $\rho_{G}^{D\Theta}$):

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

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

$$= - r_G^P\mu(G \mid d; \rho_{GS}^{D\Theta}). \tag{TA.21}$$

Since we have already verified that all of the expressions above are the same as in the base model, we can conclude that, in equilibrium, the D of type GS will mix between accepting and rejecting the plea offer, rejecting it with exactly the same probability as before. That is, the D of type GS rejects the plea offer with probability $\rho_{GS}^{D\Theta} = \rho_{G}^{D0}$; the computed value of ρ_{G}^{D0} is given in equation (TA.17).

The equilibrium rate of plea acceptance is now $\omega\lambda + [\omega + (1 - \omega)(1- \rho_{G}^{D0})](1 - \lambda)$, which is higher than in the base model wherein this rate is $(1 - \rho_{G}^{D0})(1 - \lambda)$. P is able to obtain a plea agreement with more guilty defendants, but also unavoidably sweeps up some innocent defendants as well.

The equilibrium plea offer is also affected because a defendant accepting a plea offer is no longer inferred to be guilty for sure. The plea offer in the base model is $S_b(\rho_{G}^{D0}) = \pi_T^D(G; \rho_{G}^{D0}) - r^D$, whereas the new plea offer is: $S_b(\rho_{G}^{D0}) = \pi_T^D(GS; \rho_{G}^{D0}) - r^D\mu(G \mid b; \rho_{G}^{D0})$. Note that $\pi_T^D(G; \rho_{G}^{D0})$ and $\pi_T^D(GS; \rho_{G}^{D0})$ are the same function, as all guilty defendants are type GS in the base model. So the plea offer is higher in the model with weak types.

The Scottish Verdict: Three-Outcome Regime

We now return to the base model of Sections 2 and 3, in order to investigate the Scottish Verdict. The three outcomes are "guilty" (denoted g), "not guilty" (denoted ng) and "not proven" (denoted np). The relevant threshold for a finding of g is denoted γ_g (this is assumed to be the same as the threshold γ_c for conviction in the two-outcome regime), so $1 - F_t(\gamma_g)$, $t \in \{I, G\}$, is the probability that the D of type t is found guilty. Similarly, the threshold for a finding of ng is denoted γ_{ng}, so $F_t(\gamma_{ng})$, $t \in \{I, G\}$, is the probability that the D of type t is found not guilty. Finally, the probability the D of type t receives a verdict of not proven is given by $\Delta_t \equiv F_t(\gamma_g) - F_t(\gamma_{ng})$, $t \in \{I, G\}$.

For the three-outcome regime, we assume the Strict Monotone Likelihood Ratio Property (SMLRP). That is, $f(e \mid G)/f(e \mid I)$ is strictly increasing in e for $e \in (0, 1)$. The assumption of SMLRP implies Strict First-Order Stochastic Dominance; that, in the two-outcome regime, $F_G(e) < F_I(e)$ for all $e \in (0, 1)$. Evaluating at $e = \gamma_g$ yields $F_G(\gamma_g) < F_I(\gamma_g)$; that is, an innocent D is more likely to be acquitted at trial than a guilty D. SMLRP further implies the following relationships that will be used in the three-outcome regime.

Strict Reverse Hazard Rate Dominance (SRHRD): $f_G(e)/F_G(e) > f_I(e)/F_I(e)$ for all $e \in (0, 1)$.

Strict Hazard Rate Dominance (SHRD): $f_G(e)/[1 - F_G(e)] < f_I(e)/[1 - F_I(e)]$ for all $e \in (0, 1)$.

The effect of dividing the former "acquittal" evidence interval into two sub-intervals corresponding to "not proven" and "not guilty" is to change the D of type t's payoff function from trial to the following form:

$$\pi^D_T(t) = S_c(1 - F_t(\gamma_g)) + k^D + r^D\mu(G \mid g)(1 - F_t(\gamma_g)) + r^D\mu(G \mid np)\Delta_t + r^D\mu(G \mid ng)F_t(\gamma_{ng}). \quad \text{(TA.22)}$$

Since $\gamma_g = \gamma_c$, the effect is basically to replace the expression $r^D\mu(G \mid a)F_t(\gamma_c)$ with $r^D\mu(G \mid np)\Delta_t + r^D\mu(G \mid ng)F_t(\gamma_{ng})$.

For arbitrary mixing probabilities (ρ^D_I, ρ^D_G), the beliefs are now:

$$\mu(G \mid g) = \rho^D_G(1 - \lambda)(1 - F_G(\gamma_g))/[\rho^D_G(1 - \lambda)(1 - F_G(\gamma_g)) + \rho^D_I\lambda(1 - F_I(\gamma_g))]; \quad \text{(TA.23a)}$$

$$\mu(G \mid np) = \rho^D_G(1 - \lambda)\Delta_G/[\rho^D_G(1 - \lambda)\Delta_G + \rho^D_I\lambda\Delta_I]; \quad \text{(TA.23b)}$$

$$\mu(G \mid ng) = \rho_G^D(1 - \lambda)F_G(\gamma_{ng})/[\rho_G^D(1 - \lambda)F_G(\gamma_{ng}) + \rho_I^D\lambda F_I(\gamma_{ng})]. \quad \text{(TA.23c)}$$

A sufficient condition for $\pi_T^D(G) > \pi_T^D(I)$ is that $\mu(G \mid ng) \leq \mu(G \mid np) \leq \mu(G \mid g)$. First, notice that $\mu(G \mid np) \geq \mu(G \mid ng)$ if and only if $F_I(\gamma_{ng})/F_G(\gamma_{ng}) \geq \Delta_I/\Delta_G = [F_I(\gamma_g) - F_I(\gamma_{ng})]/[F_G(\gamma_g) - F_G(\gamma_{ng})]$ or, equivalently, if and only if $F_I(\gamma_{ng})/F_G(\gamma_{ng}) \geq F_I(\gamma_g)/F_G(\gamma_g)$. These expressions are equal at $\gamma_{ng} = \gamma_g$, and SMLRP (SRHRD) implies that the ratio $F_I(e)/F_G(e)$ is strictly decreasing in e. Thus, $F_I(\gamma_{ng})/F_G(\gamma_{ng}) > F_I(\gamma_g)/F_G(\gamma_g)$ for all $\gamma_{ng} < \gamma_g$. Next, notice that $\mu(G \mid g) \geq \mu(G \mid np)$ if and only if $\Delta_I/\Delta_G = [F_I(\gamma_g) - F_I(\gamma_{ng})]/[F_G(\gamma_g) - F_G(\gamma_{ng})] \geq [1 - F_I(\gamma_g)]/[1 - F_G(\gamma_g)]$ or, equivalently, if and only if $[1 - F_I(\gamma_{ng})]/[1 - F_G(\gamma_{ng})] \geq [1 - F_I(\gamma_g)]/[1 - F_G(\gamma_g)]$. These expressions are equal at $\gamma_{ng} = \gamma_g$, and SMLRP (SHRD) implies that the ratio $[1 - F_I(e)]/[1 - F_G(e)]$ is strictly decreasing in e. Thus, $[1 - F_I(\gamma_{ng})]/[1 - F_G(\gamma_{ng})] > [1 - F_I(\gamma_g)]/[1 - F_G(\gamma_g)]$ for all $\gamma_{ng} < \gamma_g$. We therefore conclude that $\mu(G \mid ng) < \mu(G \mid np) < \mu(G \mid g)$ and thus $\pi_T^D(G) > \pi_T^D(I)$.

Since type G expects a worse outcome at trial than does type I, the equilibrium will be of the same form as before; that is, all Ds of type I will go to trial, along with a fraction of Ds of type G, denoted ρ_G^D. The plea offer will make a D of type G indifferent about accepting the plea deal and going to trial. We will again select the lowest value of ρ_G^D consistent with incentivizing P to go to trial rather than dropping the case following a rejected plea offer. Incorporating Θ's beliefs and P's beliefs (which are given by $\nu(G \mid R; \rho_G^{D\Theta}) = \rho_G^{D\Theta}(1 - \lambda)/[\rho_G^{D\Theta}(1 - \lambda) + \lambda]$), we can write P's expected payoff from trial as follows:

$$\begin{aligned}
&\nu(G \mid R; \rho_G^{D\Theta})\{S_c(1 - F_G(\gamma_g)) - k^P - r_I^P\mu(I \mid g; \rho_G^{D\Theta})(1 - F_G(\gamma_g)) - r_G^P\mu(G \mid np; \rho_G^{D\Theta})\Delta_G \\
&\qquad - r_G^P\mu(G \mid ng; \rho_G^{D\Theta})F_G(\gamma_{ng})\} \\
&+ \nu(I \mid R; \rho_G^{D\Theta})\{S_c(1 - F_I(\gamma_g)) - k^P - r_I^P\mu(I \mid g; \rho_G^{D\Theta})(1 - F_I(\gamma_g)) - r_G^P\mu(G \mid np; \rho_G^{D\Theta})\Delta_I \\
&\qquad - r_G^P\mu(G \mid ng; \rho_G^{D\Theta})F_I(\gamma_{ng})\}. \qquad \text{(TA.24)}
\end{aligned}$$

Substituting for the beliefs and collecting terms yields:

$$\mathrm{num}(\pi_T^P(\rho_G^{D\Theta})) = S_c[\rho_G^{D\Theta}(1 - \lambda)(1 - F_G(\gamma_g)) + \lambda(1 - F_I(\gamma_g))] - r_I^P\lambda(1 - F_I(\gamma_g))$$

$$- k^P[\rho_G^{D\Theta}(1 - \lambda) + \lambda] - r_G^P\rho_G^{D\Theta}(1 - \lambda)F_G(\gamma_g), \qquad \text{(TA.25)}$$

and $\mathrm{denom}(\pi_T^P(\rho_G^{D\Theta})) = [\rho_G^{D\Theta}(1 - \lambda) + \lambda]$. Notice that P's expected payoff from trial is independent of the fact that the acquittal interval has been subdivided into intervals pertaining to outcomes of not proven and not guilty. Moreover, since P's payoff from dropping the case is still $= - r_G^P\mu(G \mid d; \rho_G^{D\Theta})$, it follows that the mixing probability for the D of type G that just makes P indifferent between trial and dropping the case is exactly the same as in the two-verdict case:

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

Although the form of the equilibrium plea offer is still the same, $S_b(\rho_G^{D0}) = \pi_T^D(G; \rho_G^{D0}) - r^D$, recall that the function $\pi_T^D(G; \rho_G^{D0})$ in the three-outcome regime replaces the expression $r^D\mu(G \mid a)F_G(\gamma_c)$ with $r^D\mu(G \mid np)\Delta_G + r^D\mu(G \mid ng)F_G(\gamma_{ng})$. Because the beliefs are, in both regimes, evaluated at the same value of ρ_G^{D0} (and because $\gamma_g = \gamma_c$), we only need to compare $\mu(G \mid np)\Delta_G + \mu(G \mid ng)F_G(\gamma_{ng})$ with $\mu(G \mid a)F_G(\gamma_g)$ in order to determine whether the equilibrium plea offer is higher or lower under the three-outcome regime. It will be useful to write:

$$\mu(G \mid a) = \rho_G^{D0}(1 - \lambda)F_G(\gamma_g)/A, \text{ where } A \equiv [\rho_G^{D0}(1 - \lambda)F_G(\gamma_g) + \lambda F_I(\gamma_g)];$$

$$\mu(G \mid ng) = \rho_G^{D0}(1 - \lambda)F_G(\gamma_{ng})/B, \text{ where } B \equiv [\rho_G^{D0}(1 - \lambda)F_G(\gamma_{ng}) + \lambda F_I(\gamma_{ng})]; \text{ and}$$

$$\mu(G \mid np) = \rho_G^{D0}(1 - \lambda)\Delta_G/C, \text{ where } C \equiv [\rho_G^{D0}(1 - \lambda)\Delta_G + \lambda\Delta_I].$$

Then $\mu(G \mid np)\Delta_G + \mu(G \mid ng)F_G(\gamma_{ng}) > \mu(G \mid a)F_G(\gamma_g)$ if and only if $[(\Delta_G)^2/C] + [(F_G(\gamma_{ng}))^2/B] > [(F_G(\gamma_g))^2/A]$, which holds if and only if $[F_G(\gamma_g)B - F_G(\gamma_{ng})A]^2 > 0$. The term in brackets is nonzero because the ratio $F_G(e)/[\rho_G^{D0}(1 - \lambda)F_G(e) + \lambda F_I(e)]$ is increasing by SRHRD. Thus, the D of type G faces a higher expected punishment at trial under the three-outcome regime than under the two-outcome regime, and this also results in P making a higher plea offer in equilibrium.

On the other hand, a D of type I will face a lower expected punishment at trial under the three-outcome regime than under the two-outcome regime if $\mu(G \mid np)\Delta_I + \mu(G \mid ng)F_I(\gamma_{ng}) < \mu(G \mid a)F_I(\gamma_g)$ or, equivalently, if and only if $(1 - \mu(I \mid np))\Delta_I + (1 - \mu(I \mid ng))F_I(\gamma_{ng}) < (1 - \mu(I \mid a))F_I(\gamma_g)$. This inequality holds if and only if $[(\Delta_I)^2/C] + [(F_I(\gamma_{ng}))^2/B] > [(F_I(\gamma_g))^2/A]$, which holds if and only if $[F_I(\gamma_g)B - F_I(\gamma_{ng})A]^2 > 0$. The term in brackets is nonzero because the ratio $F_I(e)/[\rho_G^{D0}(1 - \lambda)F_G(e) + \lambda F_I(e)]$ is strictly decreasing by SRHRD.

Both Conditions 1 and 2 are easier to fulfill in the three-outcome regime. This is because the right-hand-side of Condition 1 becomes:

$$S_c(1 - F_G) + k^D]/[1 - \mu(G \mid g; \rho_G^{D0})(1 - F_G(\gamma_g)) - \mu(G \mid np; \rho_G^{D0})\Delta_G - \mu(G \mid ng; \rho_G^{D0})F_G(\gamma_{ng})],$$

and we have just shown that this denominator is smaller than the corresponding expression under the two-verdict regime. The denominator in the right-hand-side of Condition 2 is the same as in the right-hand-side of Condition 1, and (as argued above) this has become smaller with the addition of the third outcome. The numerator in the right-hand-side of Condition 2 is now:

$$k^P + k^D + r_I^P\mu(I \mid g; \rho_G^{D0})(1 - F_G(\gamma_g)) + r_G^P\mu(G \mid np; \rho_G^{D0})\Delta_G + r_G^P\mu(G \mid ng; \rho_G^{D0})F_G(\gamma_{ng}),$$

which is larger than in the two-outcome regime. Thus, both Conditions 1 and 2 hold for larger ranges of the parameter r^D.

Finally, the outside observers' expected loss from misclassification under the Scottish verdict, denoted as $\widetilde{M}(\rho_G^D)$, is lower than under the two-outcome regime. In the following expression, Θ's beliefs are as in equations (TA.23a)-(TA.23c) with $\rho_I^D = 1$.

$$\widetilde{M}(\rho_G^D) = \lambda(1 - F_I(\gamma_g))r^D\mu(G \mid g) + \lambda r^D[\mu(G \mid np)\Delta_I + \mu(G \mid ng)F_I(\gamma_{ng})] \text{ (this term is lower)}$$
$$+ \rho_G^D(1 - \lambda)(1 - F_G(\gamma_g))[r^D - r^D\mu(G \mid g)] \text{ (this term is the same)}$$

$+ \rho_G^D(1 - \lambda)\{\Delta_G[r^D - r^D\mu(G \mid np)] + F_G(\gamma_{ng})[r^D - r^D\mu(G \mid ng)]\}$ (this term is lower)

$+ \lambda(1 - F_I(\gamma_g))[r_I^P - r_I^P\mu(I \mid g)] + \lambda r_G^P[\mu(G \mid np)\Delta_I + \mu(G \mid ng)F_I(\gamma_{ng})]$ (this term is lower)

$+ \rho_G^D(1 - \lambda)(1 - F_G(\gamma_g))[r_I^P\mu(I \mid g)]$ (this term is the same)

$+ \rho_G^D(1 - \lambda)\{\Delta_G[r_G^P - r_G^P\mu(G \mid np)] + F_G(\gamma_{ng})[r_G^P - r_G^P\mu(G \mid ng)]\}$. (this term is lower)

As the equilibrium plea rejection rate is the same in both regimes, we have that $\widetilde{M}(\rho_G^{D0}) < M(\rho_G^{D0})$.

To summarize, we find that the form of the equilibrium is substantially the same under both regimes. The G-type prefers the two-outcome regime, whereas the I-type prefers the three-outcome regime. P prefers the three-outcome regime, as she obtains the same expected payoff from trial, whereas the plea offer is higher in the three-outcome regime and is accepted with the same probability. Finally, the expected loss due to misclassification experienced by outside observers is lower under the three-outcome regime.