Improving Criminal Trials by Reflecting Residual Doubt: Multiple Verdicts and Plea Bargains

Ron Siegel and Bruno Strulovici*

February 9, 2016

Abstract

We propose adding a third, intermediate verdict to the two-verdict system used in criminal trials, to distinguish between convicted defendants based on the residual doubt regarding their guilt at the end of the trial. This additional verdict improves welfare without increasing wrongful convictions or the incentives to commit a crime. We also consider plea bargains, a form of intermediate verdict, and show that a properly chosen plea in a two-verdict system increases welfare relative to any multi-verdict system, and is in fact the optimal mechanism even accounting for the incentives to commit a crime. Finally, we consider how additional verdicts affect social stigma and the incentives to gather evidence.

---

*We thank Daron Acemoglu, Robert Burns, Andy Daughety, Eddie Dekel, Louis Kaplow, Fuhito Kojima, Adi Leibovitz, Paul Milgrom, Jean Tirole, Leeat Yariv for their comments and Jennifer Reiganum for her discussion at the NBER Summer Institute Law and Economics Workshop (2015). The paper benefited from the reactions of seminar participants at UC Berkeley, Seoul National University, the NBER, the World Congress of the Econometric Society, the Harvard/MIT Theory workshop, and Caltech's NDAM conference. David Rodina provided excellent research assistance. Strulovici gratefully acknowledges financial support from an NSF CAREER Award (Grant No. 1151410) and a fellowship form the Alfred P. Sloan Foundation. Siegel: Department of Economics, The Pennsylvania State University, University Park, PA 16802, rus41@psu.edu. Strulovici: Department of Economics, Northwestern University, Evanston, IL 60208, b-strulovici@northwestern.edu.

1 Introduction

Criminal trials are imperfect: innocent defendants are sometimes convicted and guilty ones are sometimes acquitted.[1] This is unavoidable, because trials cannot always eliminate all doubt regarding defendants' guilt. How this residual doubt translates into a verdict is determined by the standard for conviction. In the United States, the standard is "beyond reasonable doubt," which reflects the view that it is more important not to punish the innocent than it is to mistakenly acquit the guilty.

One way to improve trial outcomes is to reduce the residual doubt regarding defendants' guilt. Technological advances, such as DNA profiling, sometimes achieve this, but attaining absolute certainty regarding a defendant's guilt in every case is not realistic. This paper proposes a different improvement, which builds on the observation that residual doubt varies across trials. Consider, for example, a trial in which a defendant is found guilty based on a confession and an eye witness report. These pieces of evidence may establish the defendant's guilt "beyond a reasonable doubt," but because confessions and eye-witness reports are known to be unreliable to some extent, some residual doubt remains. A similar trial in which additional evidence is available, such as clear footage of the defendant committing the crime, would result in less residual doubt regarding the defendant's guilt. This variance in residual doubt across trials cannot be reflected in a two-verdict system, in which the defendant is found either guilty or not guilty.

We propose introducing a third, intermediate verdict as a possible outcome in criminal trials. This verdict will be used when the residual doubt is close to "reasonable." Intuitively, the additional verdict is a welfare-improving alternative when the judge and/or jury are torn between convicting and acquitting a defendant. In this case, an intermediate punishment reduces the welfare loss of convicting an innocent defendant or acquitting a guilty one. The possibility of an additional verdict, has been proposed in the legal literature by Bray (2005), but has received little formal analysis.[2]

[1]For example, a recent study by Gross et al. (2014) of 7,482 death row convictions from 1973 to 2004 in the United States estimates that at least 4.1% of death-row defendants have been wrongfully convicted. Given the high burden of proof required for convictions, acquittals of guilty defendants are likely to be even more frequent.

[2]Bray's proposal concerns the addition of a "not-proven" verdict to the U.S. criminal system, which does not carry any jail time, unlike the intermediate verdicts which we introduce in Section 2. Daughety and Reinganum (2015a) consider the effect of informal sanctions on defendants and prosecutors. In an extension discussed later

Punishments in criminal trials that can be viewed as "intermediate" currently arise for other reasons. The punishment for homicide, for example, may depend on whether the defendant is charged with murder or manslaughter;[3] a single crime may lead to multiple charges, and a defendant may be convicted of only a subset of them; extenuating circumstances may substantially affect the sentence associated with a conviction. Notice, however, that the variability in punishment in these cases arises because of the variability in the nature and circumstances of the crime, not because of the degree of certainty that the defendant committed the crime. To the extent that these instruments are used to reflect residual doubt, they are not designed for this and can lead to arbitrary, unfair, and suboptimal outcomes, as explained in Section 6.

A natural question is why criminal trials today do not commonly use additional verdicts. One possibility is that such verdicts would be an open admission of the system's imperfection. After all, in an ideal world, guilty defendants would be convicted and innocent ones would be acquitted, so additional verdicts would be of no value. But criminal trials are in fact imperfect, and, as we show, welfare can be increased by recognizing this fact and introducing a third verdict. Another possibility is that the introduction of an additional verdict would give rise to several concerns. One concern is that more innocent defendants would be convicted. Another is that the incentives to commit crimes would increase. A third is that the incentives to gather evidence may be diminished. A fourth is that implementing the addition would require infeasible changes to the system, or would only be beneficial if the current punishments and conviction standard are close to optimal, which may not be the case.

We show that a third verdict can be added in a way that increases welfare and addresses all these concerns. The third, intermediate verdict will be used to distinguish among defendants who would be convicted in the current system. Among those defendants, the ones for whom more doubt remains will be punished less severely than those whose guilt is more certain. We show that for any punishment in the current system and any doubt threshold exceeding the one currently used for conviction, there is a way to set the punishments above and below the threshold that increases welfare relative to the current system and does not increase the incentives to commit crimes. This guarantees that every defendant who would be acquitted in the current system would also be acquitted in the new system. In particular, no additional innocent defendants

in that paper, they consider the effect of introducing a not-proven verdict. Daughety and Reinganum (2015b) consider several implementations of the not proven verdict through defendant choice and compensation.

[3]Homicide is an exceptional crime in that it is associated with several different criminal counts.

would be convicted. If the punishment in the current system is not too inefficiently low and incentives to commit the crime are not a significant concern, which may be the case for certain crimes of passion, we obtain the stronger result that welfare can be improved without increasing the punishment relative to the current system. This guarantees not only that no additional innocent defendants are punished in the new system, but also that those who are punished are never punished more severely than in the current system.[4]

The additional verdict can be introduced into criminal trials in the United States in several ways. One possibility is to have the jury first determine whether the defendant is guilty according to the standard used in the current system. If the jury find the defendant guilty, then in a second stage the jury would further indicate whether they find the defendant guilty "beyond a reasonable doubt" or "beyond all doubt," with a lower sentence for the former. This distinction has recently been advocated in the context of capital trials (see Section 6). A second possibility is not to change the jury's current role and instead to relegate the distinction between the two degrees of guilt uncertainty to the sentencing stage. This two-step implementation is explored in Section 7. If the jury find the defendant guilty, then the judge would determine the sentencing category based on the residual doubt regarding the defendant's guilt. A third possibility is not to change the jury's or the judge's current role and instead introduce rules or guidelines (via legislation or other means) that determine the degree of residual doubt following a conviction based on the strength of evidence produced during the trial. It may also be possible to combine some of these methods or introduce additional ones. Notice that in all the methods jurors would still be given, and should follow, the current guidelines for conviction, so the set of convicted defendants would not change.[5]

A potential concern is that jurors and other agents of the criminal justice system may reduce their effort to acquire and seriously consider the evidence if an intermediate verdict is introduced. To gain a better understanding of this issue, we consider how the introduction of a third verdict

[4]Our result about the welfare-improving addition of a third verdict holds more generally: for any multi-verdict system one can add another verdict and lower the punishments in a way that increases social welfare.

[5]Jurors are currently instructed to focus only on determining the defendant's guilt and ignore the punishment carried by a conviction (Sauer, 1995). To the extent that jurors deviate from these guidelines more in the new system than in the current system, social welfare would be further improved, as long as jurors have society's best interests in mind. Section 7 discusses how jurors' incentives may be affected by the introduction of a third verdict.

affects the value of evidence in a trial. Since gathering evidence is costly, the socially optimal amount of evidence to be gathered (and jurors' incentives to fully process this evidence) depends on the verdict structure. We show that the introduction of the third verdict generally increases the value of evidence and therefore the optimal amount of evidence that should be gathered. We obtain this result both in a two-period discrete model and in a continuous-time model in which the residual doubt changes stochastically as long as evidence is gathered.

Another approach to reducing residual doubt is to induce defendants to reveal whether they are guilty. Defendants for which this is done successfully would not go through a trial, so any residual doubt regarding their guilt would be avoided. Of course, if guilty defendants are to be punished, then simply asking defendants whether they are guilty would not work. One way to induce defendants to reveal their guilt is to offer them a plea bargain, which is an admission of guilt and a lower sentence than the one associated with a conviction.

Plea bargains are an important instrument in the United States criminal justice system.[6] Because defendants choose whether to accept the plea, and guilty defendants are (presumably) more likely to be found guilty during a trial, the plea can serve as a screening device. Building on the framework of Grossman and Katz (1983), who show that guilty defendants are more willing to take the plea, we analyze the value of plea bargains relative to other verdict systems. We show that an appropriate two-verdict system with pleas dominates *any* multi-verdict system without pleas, regardless of the number of verdicts in the system, provided that the defendant's utility from being punished is independent of his guilt. In fact, we show that there is a two-verdict system with a plea that maximizes welfare among all incentive compatible mechanisms, and does not increase the incentives to commit crimes. In this optimal mechanism, the sentence associated with a guilty verdict coincides with the sentence that is optimal when one is certain that the defendant is guilty.[7]

Despite its generality, the result on the superiority of two-verdict systems with plea bargains omits several issues. When some innocent defendants are more risk averse than guilty ones, for instance, these innocent defendants may prefer to plead guilty rather than face the lottery of the trial, particularly if the sentence set for a guilty verdict is set at level meant to be optimal conditional on a convicted defendant being surely guilty. Since some innocent defendants are

[6]More than 90% of criminal cases in the United States are settled by plea bargains (Burns (2009)).

[7]The characterization of the optimal mechanism does not follow from standard results, because the mechanism design environment does not include transfers.

also convicted, that maximal sentence may be too harsh, leading some innocent defendants to accept the plea bargain. We demonstrate (Appendix D) that when the guilty sentence is set at a suboptimally high level, the two-verdict system with a plea may be dominated by a three-verdict system.[8] The result is, however, robust in other dimensions. In particular, Silva (2015) studies a general mechanism with multiple defendants whose types (guilty or innocent) may be correlated and whose sentences may depend on one another's reports, and finds that there exists an optimal confession-inducing scheme in which confessions are met with a flat sentence similar to a plea bargain.

We also consider using the additional verdict to distinguish among defendants who would be acquitted in the two-verdict system. Since these defendants are not punished in the two-verdict system, they would not be punished in the three-verdict system. But acquitted defendants may suffer from the stigma of having been tried.[9] Because this stigma is likely related to the perceived likelihood that they are in fact guilty, distinguishing among these defendants based on the residual doubt at the end of the trial may affect the stigma they face. We treat the stigma mechanism as exogenous, since it is determined by society and cannot be legislated in the same way that sentences are. Consequently, this introduction of a third verdict does not always increase welfare, in contrast to our first result, since its socially detrimental effect on acquitted defendants who are in fact guilty may outweigh the socially beneficial effect on innocent defendants. We provide conditions under which this third verdict increases welfare, as well as comparative statics.

Several countries, including Israel, Italy, and Scotland, do in fact distinguish among acquitted defendants based on the residual doubt regarding their guilt. In Scotland, for example, a conviction in a criminal trial leads to a "guilty" verdict, but an acquittal leads to either a verdict of "not guilty" or "not proven." Neither of the two acquittal verdicts carries any jail time, but the latter indicates a higher likelihood that the defendant is in fact guilty. The likelihood is, however, insufficiently high for conviction.[10]

We also consider how to optimally incorporate a third verdict without the restriction that no

[8]One may also construct examples in which an innocent defendant who overestimates the probability of being found guilty in a trial, perhaps through persuasion or intimidation, may take a plea. In this case, a three-verdict system can again dominate the two-verdict system with a plea.

[9]Economic analyses of the stigma faced by convicts are provided by Lott (1990) and Grogger (1992, 1995)

[10]This may happen, for example, if an eye-witness testimony exists, but the testimony cannot be corroborated.

additional innocent defendants be punished. We show that an optimal three-verdict system will generally punish defendants more frequently than the two-verdict system, since the intermediate verdict will carry a positive sentence, but the additional defendants who are punished, as well as some defendants who would be punished in the two-verdict system, optimally receive a lower punishment than convicted defendants in the two-verdict system. However, those defendants who are punished in the two-verdict system and regarding whose guilt little uncertainty remains at the end of the trial are optimally punished more severely in the three-verdict system.[11]

The appendix provides a micro-foundation for the Bayesian formulation used in later parts of the paper. It establishes that trial technology conceptualized as a mapping from accumulated evidence to a verdict can always be reformulated in Bayesian fashion: accumulated evidence is a signal that turns the prior probability that the defendant is guilty into a posterior probability, on which the verdict is based. Moreover, this transformation establishes a relationship between two notions of 'incriminating' and 'exculpatory' evidence. One notion is based on decisions and the other on beliefs. What makes a piece of evidence 'incriminating' is the fact that it increases the likelihood of guilt of a defendant and, hence, results in a longer expected sentence. In particular, there is no loss of generality when one says that a guilty defendant is more likely to generate incriminating evidence than an innocent defendant.

2 Reflecting residual doubt in trial outcomes

We consider a trial whose objective is to determine whether a defendant is guilty of committing a certain crime and to deliver the corresponding sentence. In our baseline model the trial is summarized by two numbers: the probability π_g that the defendant is found guilty if he is actually guilty, and the probability π_i that the defendant is found guilty if he is actually innocent.[12] Corresponding to a guilty verdict is a sentence $s > 0$, interpreted as jail time (so a

[11] Daughety and Reinganum (2015a) consider how the effect of informal sanctions on defendants and prosecutors affect the plea bargaining process and its acceptance rate, and consider the effect of introducing a not-proven verdict in this context. Daughety and Reinganum (2015b) consider two implementations of the not-proven verdict. In the first one, the defendant can choose between the standard binary verdict system and the system with a not-proven verdict. In equilibrium, all defendants choose the latter verdict. The authors also analyze an alternative implementation in which some defendants who are found not guilty are compensated.

[12] It is natural to assume that $\pi_g > \pi_i$, i.e., a defendant is more likely to be found guilty if he is actually guilty than if he is innocent. This restriction is, however, not required for this section.

higher value of s corresponds to a higher punishment).[13]

Society's goal is to avoid punishing innocent defendants and adequately punish guilty ones. This dual goal is modeled by a welfare function, denoted W. Jailing an innocent defendant for s years leads to a welfare of $W(s,i)$, with $W(0,i) = 0$ and W decreasing in s. Jailing a guilty defendant leads to a welfare of $W(s,g)$, which has a single peak at $\bar{s} > 0$. Thus, $\bar{s}$ is the punishment deemed optimal by society if it is certain that the defendant is guilty.

The relative importance of these objectives depends on the prior probability λ that the defendant is in fact guilty. The more likely the defendant is ex-ante to be guilty, the more important it is to adequately punish him if he is in fact guilty; the less likely the defendant is ex-ante to be guilty, the more important it is to avoid punishing him if he is in fact innocent. This is captured by the ex-ante social welfare from the defendant going to trial when the punishment of being found guilty is s:

$$\mathcal{W}_2(s) = \lambda \left[\pi_g W(s,g) + (1-\pi_g)W(0,g)\right] + (1-\lambda)\left[\pi_i W(s,i) + (1-\pi_i)W(0,i)\right]. \qquad (1)$$

Since $W(\cdot,i)$ is decreasing and $W(\cdot,g)$ peaks at $\bar{s}$, it is never optimal to choose $s > \bar{s}$. In what follows, we restrict attention to sentences s lying in $[0,\bar{s}]$.

2.1 Intermediate "guilty" verdict

We introduce a third verdict in such a way that those defendants who would be convicted in the two-verdict system now receive one of two "guilty verdicts," which we denote 1 and 2. Defendants who would be acquitted in the two-verdict system are still acquitted and are released. The distinction between the two "guilty" verdicts may be based on the evidence available before and during the trial, so that among the collections of evidence that would lead to a conviction in the two-verdict system some lead to verdict 1 and the remaining to verdict 2.[14] Denote by π_i^1 the probability that the defendant receives verdict 1 if he is innocent, and define π_i^2, π_g^1, and π_g^2

[13] We leave aside such issues as mitigating circumstances, which are tangential to the focus of the paper.

[14] Evidence leading to a homicide conviction in the two-verdict system may include, for example, the discovery, in the defendant's house, of the gun from which the bullet was fired, a confession by the defendant, a death threat made by the defendant to the victim shortly before the murder, or a union of any subset of these.

similarly. Because the probability of not acquitting the defendant does not change,[15] we have

$$\pi_i = \pi_i^1 + \pi_i^2 \quad \text{and} \quad \pi_g = \pi_g^1 + \pi_g^2.$$

Without loss of generality[16]

$$\frac{\pi_g^1}{\pi_i^1} < \frac{\pi_g}{\pi_i} < \frac{\pi_g^2}{\pi_i^2},$$

so verdict 1 is an "intermediate verdict:" an innocent defendant is more likely to receive verdict 1, relative to a guilty defendant, than verdict 2.

Let s_j denote the sentence associated with verdict j. Given s_1 and s_2, the expected welfare is given by

$$\begin{aligned}\mathcal{W}_3(s_1, s_2) = \lambda \left[\pi_g^1 W(s_1, g) + \pi_g^2 W(s_2, g) + (1 - \pi_g) W(0, g)\right] + \\ (1 - \lambda) \left[\pi_i^1 W(s_1, i) + \pi_i^2 W(s_2, i) + (1 - \pi_i) W(0, i)\right].\end{aligned} \tag{2}$$

Our first result shows that s_1 and s_2 can be chosen so that this welfare is higher than the one in the two-verdict system, provided that the sentence s associated with a conviction in the two-verdict system is interior, i.e., $s < \bar{s}$.

Proposition 1 *For any interior sentence s of the two-verdict system and any verdict technologies π_i, π_g, π_i^j, etc., there are sentences s_1 and s_2 such that $s_1 < s < s_2$ and $\mathcal{W}_3(s_1, s_2) > \mathcal{W}_2(s)$.*

A key aspect of Proposition 1 is that the three-verdict system does not increase the probability of punishing the innocent, compared to the two-verdict system. Instead it modifies the sentence to reflect the richer information that verdicts 1 and 2 convey regarding the relative likelihood of the defendant being guilty or innocent.[17]

Proof. First, observe that $\mathcal{W}_3(s, s) = \mathcal{W}_2(s)$: if we give verdicts 1 and 2 the sentence associated with the guilty verdict of the two-verdict case, then we clearly obtain the same welfare as in the two-verdict case. We are going to create a strict welfare improvement by slightly perturbing the

[15] In keeping with most of the literature on trial design, we take a reduced-form approach to modeling these probabilities. We provide a micro-foundation for these probabilities in Appendix C. Section 7 discusses how the explicit consideration of jurors' incentives might affect the analysis and reviews the relevant literature.

[16] For any a, b, c, d of $\mathbb{R}_{++}$ we have $\min\{a/b, c/d\} \leq (a + c)/(b + d) \leq \max\{a/b, c/d\}$, with strict inequalities if $a/b \neq c/d$, a generic condition which we will assume throughout (it is easy to impose conditions to guarantee it: for example, one can rank bodies of evidence in terms of the posterior that they generate, as in Section C).

[17] While our model abstracts for now from the incentives to commit crimes, our design can easily accommodate an increase in s_2 that maintains deterrence, as shown in the next section.

sentences s_1 and s_2. Consider any small $\varepsilon > 0$ and let $s_1 = s - \varepsilon$ and $s_2 = s + \varepsilon\gamma$. The welfare impact of this perturbation is

$$\mathcal{W}_3(s_1, s_2) = \mathcal{W}_2(s) + \lambda\varepsilon W'_g(-\pi^1_g + \gamma\pi^2_g) + (1-\lambda)\varepsilon W'_i(-\pi^1_i + \gamma\pi^2_i) + o(\varepsilon), \tag{3}$$

where W' denotes the derivative of W with respect to its first argument. Since $W(\cdot, i)$ is decreasing, $W'(s, i)$ is negative. Similarly, because $s \leq \bar{s}$ and $W(\cdot, g)$ is increasing on that domain, $W'(s, g)$ is positive. Since $\pi^1_g/\pi^2_g < \pi^1_i/\pi^2_i$, we can choose γ between these two ratios. Doing so guarantees that $-\pi^1_g + \gamma\pi^2_g$ is positive and $-\pi^1_i + \gamma\pi^2_i$ is negative, which shows the claim. ■

While the improvement in Proposition 1 does not increase the probability of punishing an innocent defendant (or a guilty one), an erroneously convicted defendant may face a worse sentence ex-post, because $s_2 > s$. The next next result shows that if the sentence associated with a conviction in the two-verdict system is interior and optimal, then there is an improvement that does not increases the sentence.

Proposition 2 *Suppose that s^* maximizes $\mathcal{W}_2(s)$ and is interior. Then, there exists $s_1 < s$ such that $\mathcal{W}_3(s_1, s^*) > \mathcal{W}_2(s^*)$.*

The proof of Proposition 2 shows that the result holds even when the original sentence was not optimal, as long as it was not too suboptimally low. Thus, it may be generally possible to improve upon the two-verdict system even under the strong restriction of not harming any innocent defendant more than in the two-verdict system.

Proof. By construction s^* maximizes

$$\lambda\left[\pi_g W(s, g) + (1-\pi_g)W(0, g)\right] + (1-\lambda)\left[\pi_i W(s, i) + (1-\pi_i)W(0, i)\right]$$

with respect to s. Since s^* is interior, it must satisfy the first-order condition

$$\lambda\pi_g W'(s^*, g) + (1-\lambda)\pi_i W'(s^*, i) = 0. \tag{4}$$

Now consider the derivative of $\mathcal{W}_3(s_1, s^*)$ with respect to s_1, evaluated at $s_1 = s^*$. From (3), we have

$$\left.\frac{\partial \mathcal{W}_3(s_1, s^*)}{\partial s_1}\right|_{s_1 = s^*} = \lambda\pi^1_g W'(s^*, g) + (1-\lambda)\pi^1_i W'(s^*, i). \tag{5}$$

Since $\frac{\pi^1_g}{\pi^1_i} < \frac{\pi_g}{\pi_i}$, $W'(s^*, g) > 0$ and $W'(s^*, i) < 0$, the first-order condition (4) implies that the right-hand side of (5) is strictly negative. This shows that decreasing s_1 below s^* strictly improves welfare, yielding the desired improvement. ■

2.2 Incentives to commit a crime

Our analysis so far was conducted from the point at which the defendant was apprehended and brought to trial, and we considered the effect that changing the trial system has on society's welfare with respect to this defendant. An important aspect from which we have abstracted is the incentives to commit the crime in the first place. These incentives play a key role in seminal economic analyses of criminal justice systems (Becker (1966), Stigler (1970)) and received renewed emphasis from Kaplow (2011). The incentives to commit a crime may *a priori* be influenced by the introduction of a third verdict. We show that many of our welfare results continue to hold even when these incentives are taken into account.

For this, suppose that society's overall welfare can be written as a function $T(CW; d; c)$, where CW is the court welfare, i.e., the welfare we have considered in the paper up to this point ($\mathcal{W}_2$ or $\mathcal{W}_3$), d is the fraction of the population that commits a crime, and c is the direct social cost of the crime. It is reasonable to assume that T increases in CW. Since the changes to the trial system we consider increase CW, in order to show that they increase overall welfare T it suffices to show that they can be introduced in a way that does not change d. For this, consider an individual's incentives to commit a crime. In deciding whether to commit the crime, the individual weighs the direct benefit he obtains from the crime (which may vary across individuals) against his expected cost of committing the crime, which is the probability that he will face the trial system, i.e., arrested with enough evidence to justify criminal proceedings, times his expected (dis)utility from going through the trial system.

Thus, to show that d does not change, it is enough to show that the changes we propose do not affect the expected utility that an individual who commits the crime obtains from going through the trial system.[18] Consider first the introduction of an additional verdict in Proposition 1. The key for this result was choosing a ratio γ of the increase in the sentence associated with a higher degree of guilt to the decrease in the sentence associated with a lower degree of guilt. The range of welfare-improving ratios is $\left[\pi_g^1/\pi_g^2, \pi_i^1/\pi_i^2\right]$, which is independent of the function $W(\cdot, g)$. Replacing $W(\cdot, g)$ by the individual's utility function, $u(\cdot)$, at the sentencing stage, and setting $\gamma = \pi_g^1/\pi_g^2$ would, to a first order, make a guilty defendant indifferent between the two schemes. Choosing $\gamma = \pi_g^1/\pi_g^2$ therefore increases CW without changing d. This γ ratio works regardless

[18] We make the reasonable assumption that the probability that an individual who commits a crime will face the trial system does not change if d does not change.

of the sentence used in the two-verdict system. The result thus applies regardless of whether the original sentence was determined by considering the incentives to commit the crime.

In contrast to Proposition 1, Proposition 2 does not generally hold when incentives to commit the crime are taken into account.[19] One should note, however, that when the two-verdict sentence is optimized to account for the incentives to commit crime, this makes Proposition 2 more likely to hold than when the two-verdict sentence is chosen to maximize $\mathcal{W}_2$. Indeed, when crime incentives are taken into account in the social objective, the optimal sentence is higher than the sentence maximizing $\mathcal{W}_2$ since, on the margin, a higher sentence reduces crime incentives. Since the sentence is, from an interim perspective—i.e., once the defendant is brought to trial—too high, lowering it for the intermediate verdict increases interim social welfare more than the same decrease for the sentence that maximizes $\mathcal{W}_2$.

2.3 The Bayesian conviction model

The analysis thus far has not imposed any structure on how verdicts were determined. Because some later parts of the paper will require it, we now show how to specialize the setting to a class of verdicts based on the posterior probability that the defendant is guilty. Starting with a prior probability λ, the trial generates evidence that is used to form the posterior. This is summarized by distributions $F(\cdot|g)$ and $F(\cdot|i)$, which describe the posterior based on whether the defendant is actually guilty or innocent.[20] For expositional convenience, we assume that $F(\cdot|g)$ and $F(\cdot|i)$ have positive densities $f(\cdot|g)$ and $f(\cdot|i)$.

In a two-verdict system based on the defendant's posterior, it is natural to follow a cut-off rule. Appendix C shows that any "reasonable" verdict rule based on evidence in the two-verdict system can be formalized as a Bayesian model with posterior cut-off rule. If the posterior p is below a threshold p^*, then the defendant is acquitted, receiving a sentence of $s = 0$. If instead p exceeds p^*, then the defendant receives a sentence $s^* > 0$. The cutoff rule is a particular case

[19] This is hardly surprising, because lowering the sentence for one verdict without increasing it for the other verdict leads to a lower expected disutility for a guilty defendant, and this may lead to more individuals committing crimes.

[20] In order to match the prior λ, the distributions must satisfy the conservation equation

$$\lambda = E[p] = \lambda \int_0^1 p dF(p|g) + (1-\lambda) \int_0^1 p dF(p|i).$$

of the previous section, with $\pi_g = Pr[p > p^*|g] = 1 - F(p^*|g)$ and $\pi_i = 1 - F(p^*|i)$.

The ex-ante social welfare is given by

$$\begin{aligned}\mathcal{W}_2(p^*, s^*) = \lambda \left[(1 - F(p^*|g))W(s^*, g) + F(p^*|g))W(0, g)\right] + \\ (1 - \lambda) \left[(1 - F(p^*|i))W(s^*, i) + F(p^*|i)W(0, i)\right].\end{aligned} \tag{6}$$

In what follows, we will denote by (p^*, s^*) the cutoff and sentence used in the two-verdict system. These variables may be chosen to maximize (1). In that case, they correspond to the utilitarian optimum for the two-verdict case.

2.4 Multi-verdict systems

Our analysis can be extended to more than three verdicts, and doing so prepares the ground for the general optimality result, in Section 3, concerning plea bargains. Granted an arbitrary number of verdicts, one would wish to associate to each posterior belief p the sentence $s(p)$ maximizing the welfare objective

$$pW(s, g) + (1 - p)W(s, i) \tag{7}$$

with respect to s. Since both $W(\cdot, g)$ and $W(\cdot, i)$ are decreasing beyond the ideal punishment $\bar{s}$ for a guilty defendant, any optimizer of (7) is lower than $\bar{s}$. Moreover, rewriting the objective function as

$$\mathcal{W}(p, s) = p[W(s, g) - W(s, i)] + W(s, i),$$

we notice that it is supermodular in (p, s).[21] This implies that the selection of maximizers of (7) is isotone. In particular, there exists a nondecreasing selection $s(p)$ of optimal sentences.

The arguments used for Propositions 1 and 2 easily generalize to yield the following results. For $k \geq 2$, we define a k-verdict system by a vector $(p_0, s_0, p_1, s_1, \ldots, p_{k-1}, s_{k-1})$ of strictly increasing cutoffs and sentences, with $p_0 = 0$, $p_{k-1} < 1$, $s_0 = 0$ and $s_{k-1} \leq \bar{s}$. In this system, a defendant gets sentence $s_{k'}$ whenever his posterior p lies in $(p_{k'}, p_{k'+1})$.

Proposition 3 *Suppose that the signal distributions are continuous for both the guilty and innocent defendants. Then, for any k-verdict system there is a $k+1$ verdict system that strictly*

[21] $W(s, g)$ increases in s over the relevant range $[0, \bar{s}]$ while $W(s, i)$ is decreasing in s. This implies that $\partial\mathcal{W}/\partial p = W(s, g) - W(s, i)$ increases in s and, hence, supermodularity of $\mathcal{W}(p, s)$. See Milgrom and Shannon (1994).

increases welfare. Moreover, if a k-verdict system is optimal among all k-verdict systems and either $k > 2$ or $k = 2$ and $s_1 < \bar{s}$, then there is a $k+1$-verdict system that strictly improves upon it and has lower sentences.

2.5 Welfare maximization with three verdicts

Although normatively appealing, the cutoff and sentence restrictions limit the welfare improvement that can be attained, and it is natural to ask what the optimal three-verdict system looks like. The result is provided by the following proposition.

Suppose that (p^*, s^*) are optimal in the two-verdict system, and let $(p_1^*, p_2^*, s_1^*, s_2^*)$ be optimal in the three-verdict system (if the posterior is below p_1^*, then the sentence is 0, if the posterior is between p_1^* and p_2^*, then sentence is s_1^*, etc.).

Assumption: $W(s, i)$ and $W(s, g)$ are concave and twice differentiable in s, with a strictly negative second derivative, and the posterior distributions $F(s|i)$ and $F(s|g)$ are both absolutely continuous in s.

Proposition 4 *$p_1^* \leq p^* \leq p_2^*$ and $s_1^* \leq s^* \leq s_2^*$.*

Intuitively, the optimal sentence reflects the likelihood that the agent is guilty. Thus, 'higher' sets of priors will lead to a longer sentence. This intuition, however, only explains the second part of Proposition 4; it does not explain why the optimal three-verdict cutoffs lie on both sides of the optimal two-verdict cutoff. The proof of this proposition requires several steps explained in Appendix A.

3 Plea bargaining

More than 90% of criminal cases in the United States conclude in a plea bargain instead of a trial. Plea bargains can be viewed a kind of third verdict, which corresponds to an intermediate sentence that is lower than the one associated with a trial conviction. This third verdict is different from what has been discussed so far, because it involves a strategic decision by the defendant of whether to take the plea, in contrast to his passive role in a multi-verdict trial. As we shall see, this strategic aspect has a substantial impact on welfare.

We model pleas similarly to Grossman and Katz (1983)—hereafter "GK." In the first stage, the defendant is offered a plea sentence, denoted s^b. If the defendant accepts the plea, he gets

this sentence and the case is concluded. If he rejects the plea, he goes to trial and faces the same signal structure as in the previous sections. The welfare functions $W(\cdot, i)$ and $W(\cdot, g)$ are also as in the previous sections.

GK show that the optimal system with a plea bargain is separating: the plea s^b is chosen so to make a guilty defendant indifferent between taking the plea and going to trial, a guilty defendant takes the plea, and an innocent defendant goes to trial.

We now show that an *any* multi-verdict system without pleas, no matter how many verdicts it has, is dominated by separating plea bargain system with only two verdicts. In fact, we will show that such a the plea system with two verdicts is optimal within a much broader class of mechanisms.

3.1 The welfare value of plea bargaining

We denote by $t \in T$ the signal (evidence) generated during the trial. We assume that t is real-valued and let $F_g(t)$ and $F_i(t)$ respectively denote the signal distributions conditional on the defendant being guilty or innocent.[22] We assume that these distributions are absolutely continuous with positive densities $f_g(t)$ and $f_i(t)$. We also assume, without loss of generality, that the signal space is $T = [0, 1]$ and that the signals are ordered according to the monotone likelihood ratio property (MLRP): the density ratio $f_g(t)/f_i(t)$ is increasing in t (see Appendix C.1).

A (measurable) multi-verdict system is a map $s : t \to s(t)$ from signals into sentences. We assume that $s(t)$ lies in $[0, \bar{s}]$ for all t, where $\bar{s}$ is the ideal sentence for a surely guilty defendant.

Proposition 5 *For any multi-verdict system $s(\cdot)$, there exists a two-verdict system with a plea that generates higher welfare.*

Proof. We begin by constructing a two-verdict system $\hat{s}$ that give the guilty defendant the same expected utility as $s(\cdot)$. In this system, there is a cutoff $\hat{t}$ below which the sentence is zero and above which the sentence is $s^M = \max_{t\in[0,1]} s(t)$. Moreover, the cutoff is chosen so that

$$U^g = \int_0^1 u(s(t))f_g(t)dt = \int_0^1 u(\hat{s}(t))f_g(t)dt = u(0)F_g([0,\hat{t}]) + u(s^M)F_g([\hat{t},1]) = \hat{U}^g, \qquad (8)$$

[22]General evidence structures are discussed in Appendix C.1. If signals were multidimensional, we could order them according to their likelihood ratios and treat the resulting ratio as the signal, so that the real-valued assumption is without loss as long as the likelihood ratio of each signal is well-defined. For example, if T is a Borel subset of $\mathbb{R}^K$ for some dimension K, the ratios will be well defined as long as the signal distributions are absolutely continuous with respect to the Lebesgue measure induced over T and have positive densities.

recalling that $u(s)$ denotes the defendant's utility from getting sentence s, and u is decreasing and concave. That such a $\hat{t}$ exists follows because the right-hand side of (8) is continuous in the cutoff t, ranging all values from $u(0)$ to $u(s^M)$, and because U^g clearly lies between $u(0)$ and $u(s^M)$ as a convex combination of utilities that lie in this interval. Moreover, the new verdict system increases the expected utility of an innocent defendant. To show this, notice that by construction we have

$$\int_0^{\tilde{t}} [u(\hat{s}(t)) - u(s(t))] f_g(t) dt \geq 0$$

for all $\tilde{t} \in [0, 1]$. Since $f_i(t)/f_g(t)$ is positive and decreasing in t, this implies that[23]

$$\int_0^1 [u(\hat{s}(t)) - u(s(t))] f_i(t) dt \geq 0,$$

or

$$\hat{U}^i \geq U^i.$$

We now introduce the plea s^b, setting it so as to make the guilty defendant indifferent between taking the plea and going to trial in the two-verdict system: that is, we choose s^b so that

$$u(s^b) = U^g = \hat{U}^g.$$

Since the guilty is indifferent, the innocent strictly prefers going to trial because i) guilty and innocent share the same utility function, but ii) an innocent defendant is less likely to be found guilty than a guilty one, so the trial is more appealing (see GK for a formal argument).

Since the innocent benefits from the new verdict system, we will have shown that this system improves on the original one if we prove that the social welfare conditional on facing the guilty defendant is also higher. This welfare is equal to $W(s^b, g)$. Because the defendant is risk averse (u is concave), s^b is greater than the average sentence $\tilde{s} = \int_0^1 s(t) f_g(t) dt$ that the guilty gets if he goes to trial. And because $W(\cdot, g)$ is concave, we have $W(\tilde{s}, g) \geq \int_0^1 W(s(t), g) f_g(t) dt$. Finally, since $s^b \geq \tilde{s}$ and $W(\cdot, g)$ is increasing, we conclude that $W(s^b, g)$ dominates the expected social welfare conditional on facing the guilty.

[23]The argument proceeds by a simple integration by parts. See Quah and Strulovici (2012, Lemma 4) for a similar proof in a more general environment. The claim may also be shown by showing that the defendant's expected utility has the single-crossing property in the defendant's type: the integrand has the single-crossing property in t and the type of the agent is affiliated with the posterior, which implies that the expected utility has the single-crossing property (see, e.g., Athey, 2002).

In conclusion, the new two-verdict system with plea improves social welfare regardless of whether the defendant is innocent or guilty. In particular, it is an improvement regardless of the prior distribution. Finally, notice that the improvement is strict if either u or $W(\cdot, g)$ is strictly concave. ■

By modifying the proof, it is possible to prove that the following, stronger result. All the verdict systems, with and without pleas, may be seen as particular mechanisms. It is well known from the mechanism design literature that in the present setting it is enough to consider direct revelation mechanisms in which it is optimal for the defendant to report his type truthfully: the defendant makes a reports $\hat{\theta}$ of his type (guilty or innocent) and is then assigned a sentence $s(t, \hat{\theta})$ that depends on his report and on the signal t generated during trial. A mechanism is feasible if $s(t, \hat{\theta}) \leq \bar{s}$ for all t and $\hat{\theta}$, i.e., it does not punish the defendant more than would be optimal if the defendant were known to be guilty. A feasible mechanism is optimal if it maximizes welfare given the prior probability λ that the defendant is guilty.

Proposition 6 *There is a unique optimal mechanism. This mechanism takes the form of a two-verdict system with a plea: $s(\cdot, g)$ is constant (i.e., like a plea), and $s(\cdot, i)$ is a two-step function, which jumps from 0 to $\bar{s}$. The incentive compatibility constraint of the guilty defendant binds. The signal cutoff at which $s(\cdot, i)$ jumps from 0 to $\bar{s}$ decreases in the prior.*

Proof. Consider a direct mechanism $s(\cdot, \cdot)$ in which it is optimal for the defendant to report his type truthfully. We begin by replacing $s(\cdot, i)$ with a two-verdict system $\hat{s}(\cdot, i)$ with a cutoff $\hat{t}$ below which the sentence is zero and above which the sentence is $\bar{s}$. The cutoff is chosen so that the innocent defendant is indifferent between $s(\cdot, i)$ and $\hat{s}(\cdot, i)$, that is,

$$U^i = \int_0^1 u(s(t,i)) f_i(t) dt = \int_0^1 u(\hat{s}(t,i)) f_i(t) dt = u(0) F_i([0, \hat{t}]) + u(\bar{s}) F_i([\hat{t}, 1]) = \hat{U}^i.$$

The guilty defendant prefers $s(\cdot, i)$ to $\hat{s}(\cdot, i)$, i.e., his incentive compatibility continues to hold, when $s(\cdot, i)$ is replaced with $\hat{s}(\cdot, i)$. This is because by construction we have

$$\int_0^1 [u(s(t,i)) - u(\hat{s}(t,i))] f_i(t) dt = 0,$$

and since $h(\cdot) = u(s(\cdot, i)) - u(\hat{s}(\cdot, i))$ crosses 0 once from below on $[0, 1]$ and $f_i(t)/f_g(t)$ is positive and decreasing in t, we obtain (see the previous footnote)

$$\int_0^1 [u(s(t,i)) - u(\hat{s}(t,i))] f_g(t) dt \geq 0.$$

Thus, because the guilty defendant prefers $s(\cdot, g)$ to $s(\cdot, i)$, he also prefers $s(\cdot, g)$ to $\hat{s}(\cdot, i)$. Now replace $s(\cdot, g)$ with the constant sentence s^b such that the guilty defendant is indifferent between s^b and $s(\cdot, g)$, that is,

$$u(s^b) = \int_0^1 u(s(t, g)) f_g(t) dt.$$

This increases welfare because the guilty defendant and society are risk averse, as in the proof of Proposition 5.

Because the guilty defendant is indifferent between s^b and $s(\cdot, g)$, he prefers s^b to $\hat{s}(\cdot, i)$. If the preference is strict, modify $\hat{s}(\cdot, i)$ by increasing $\hat{t}$ until the guilty defendant is indifferent between s^b and $\hat{s}(\cdot, i)$. This increases welfare since it increases the utility of the innocent defendant, and also guarantees that the innocent defendant prefers $\hat{s}(\cdot, i)$ to s^b (because the guilty defendant is indifferent between the two). This shows that the optimal mechanism is of the form described in the statement of the proposition, and that the incentive constraint of the guilty defendant binds. Thus, each such mechanism is pinned down by the cutoff $\hat{t}$. Finally, it is straightforward to see that the welfare-maximizing $\hat{t}$ decreases in the prior λ. ■

Proposition 6 also applies when incentives to commit the crime are taken into account.[24] That is, even when changing the trial system may affect the individuals' decision to commit the crime, a two-verdict system with a plea is still optimal. To see this, it is enough to show that given any mechanism, there exists a two-verdict system with a plea that improves welfare and does not change the set of individuals who commit a crime (in the notation of Section 2, the proportion d of individuals who commit the crime does not change). Beginning with some mechanism, each step of the proof of Proposition 6 alters the mechanism in a way that increases welfare but leaves the expected utility of a guilty defendant unchanged. Thus, the two-verdict system with a plea that improves upon the original mechanism increases welfare and generates the same expected utility for a guilty defendant that the original mechanism did, so d is unchanged.

Despite these results, pleas have been severely criticized for leading innocent defendants to

[24]Becker (1966) already noted the optimality of an extreme punishment, using a different argument: to achieve a given level of deterrence, using a higher level punishment allows society to spend less effort on catching and prosecuting criminals while keeping the expected punishment (or expected disutility of punishment) unchanged. Here, by contrast, using a higher level of punishment in a trial conviction helps relax the incentive compatibility constraint of the guilty defendant.

accept jail time rather than go to trial. This may result from the fact that sentences given at trial are excessively harsh, which is a problem that has been pointed out repeatedly.[25] Section D provides an example that illustrates this idea. It should be noted, however, that many of the criticisms leveled at plea bargaining can, at least in principle, be addressed. In the United States, a defendant is entitled to competent counsel at the plea bargaining stage in all federal trials as well as in some state-level trials.

4 Value of evidence with a third verdict

The previous sections have taken as given the technology that generates evidence in favor of or against the defendant. Gathering evidence is costly, however, and the amount of evidence generated in a case depends on the incentives of the agents involved in the evidence-gathering process: law enforcement officers, prosecutors, experts, etc.

Leaving aside the possible biases in these agents' behavior, the socially optimal amount of evidence to be gathered in a case clearly depends on the verdict structure. For example, a trial system in which a single verdict is given regardless of the evidence produced clearly eliminates any value of gathering evidence. This dependency has led to a criticism of plea bargaining: that so many defendants take pleas reduces incentives for evidence gathering.

This section compares the impact on evidence gathering of introducing a third verdict. For simplicity, we focus on the setting of Section 2.1 with the Bayesian conviction model.

A (possibly multi-) verdict system leads to welfare

$$w(p) = pW(s(p), g) + (1-p)W(s(p), i), \tag{9}$$

where $p \mapsto s(p)$ is a step function that starts at zero, has two levels in a two-verdict system, and three levels in a three-verdict system. The welfare function $w(p)$ is piecewise linear. It start at 0, and decreases until a kink at which the sentence jumps from 0 to a positive level. Figure 1 represents the welfare function for the optimal two-verdict system when $W(\cdot, g)$ and $W(\cdot, i)$ are quadratic, for parameters given in the appendix.

The kink occurs at the cutoff $p^* = 1/3$, at which the sentence jumps from 0 to 2/3. Figure 2 represents the welfare function for the optimal three-verdict system obtained by adding an

[25] See for example Judge Rakoff's "Why Innocents Plead Guilty," in the *The New York Review*, (November 20, 2014) and Justice Kagan's opinion in Supreme Court Ruling No. 13-7451 on Yates vs. U.S.

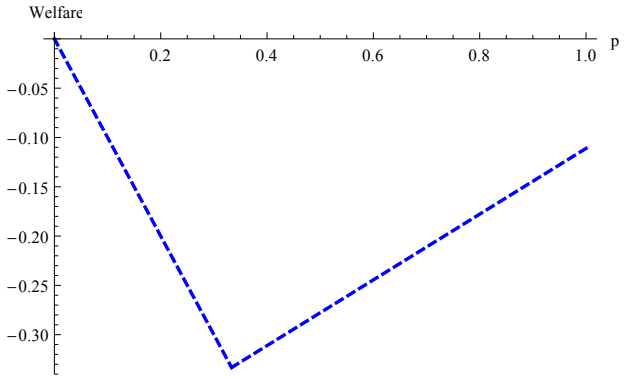

Figure 1: Welfare function, 2 verdicts.

intermediate verdict and keeping the highest sentence at 2/3. The first cut-off is $p_1 = p^* = 1/3$, and the second cut-off is $p_2 = 1/2$. The welfare function is discontinuous at p_1: this reflects the fact that p_1 is not chosen optimally, but is rather "inherited" from the two-verdict system. In contrast, because p_2 is chosen optimally, the welfare function is kinked but continuous at p_2.

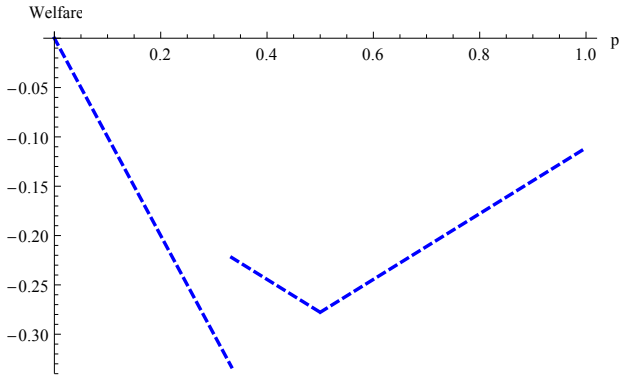

Figure 2: Welfare function, 3 verdicts.

Actual evidence formation processes are complex, involving various actors of different types – forensic experts, lawyers, witnesses—and different forms of evidence. To model evidence formation, we must abstract from much of this complexity. Instead, we take the viewpoint of a social planner who may gather information until a verdict is reached.

The tradeoff at the heart of this task is clear: more effort spent gathering evidence means higher costs for society but more precise information about the defendant's guilt. We discuss two ways to model this tradeoff (there are, of course, many others). This first is a one-shot evidence-gathering decision, which already captures the rough intuition for why two-verdict and three-verdict systems differ in their effects on evidence gathering. The second is a continuous evidence-gathering process, which provides a more visually appealing representation of the impact of a

third verdict on evidence gathering.

4.1 One-shot evidence gathering

Suppose the planner decides whether to gather evidence, which has a cost $c > 0$. Starting with a prior p_0, the evidence returns a higher probability of guilt, say $p_0 + \Delta$ with probability $1/2$, and a lower probability $p_0 - \Delta$ also with probability $1/2$. The belief process is a martingale: the mean of the posterior p' is equal to the prior: $\frac{1}{2}(p + \Delta) + \frac{1}{2}(p - \Delta) = p$.

When is evidence gathering socially desirable? Suppose first that the prior is close to 0, so that the posterior p' surely lies below the cutoff p_1. Then, the additional evidence has no value as the defendant will be acquitted in all cases. Similarly, if p_0 is high enough for p' to lie above the cutoff p_1 no matter what, the additional evidence have no value as the defendant will be convicted regardless of p'.

Consider now the case of three verdicts. For p_0 slightly below p_1 and Δ such that $p_0 + \Delta$ lies above p_1, the value of evidence is higher than in the two-verdict case because a positive belief update triggers a large improvement in welfare (see Figure 2). For p in a neighborhood of p_2, the value of evidence is also positive due to the convex kink there, whereas it is 0 (for Δ small enough) in the two-verdict case.

For p_0 slightly above p_1 however, additional evidence may be more valuable in the two-verdict case, which creates a "doughnut hole:" additional evidence is more valuable in the three-verdict case than in the two-verdict case for more extreme beliefs, and less valuable in some intermediate region. This result is easier to visualize in the next model, where evidence gathering is more gradual.

4.2 Continuous evidence gathering

Now suppose that evidence is gathered for continuously. As long as evidence is gathered, a flow cost of c is incurred. During this time the belief p_t that the defendant is guilty evolves as a martingale according to a continuous signal, modeled as in Bolton and Harris (1999):

$$dp_t = Dp_t(1 - p_t)dB_t,$$

where B is the standard Brownian motion and D is a measure of the quality of the signal: the higher D is, the faster p evolves toward the true probability that the defendant is guilty (0 or

1). At some time T, the evidence formation process is stopped and the verdict is chosen based on the posterior p_T, which results in social welfare $w(p_T)$.

Let $v(p)$ denote the value function corresponding to stopping optimally Adapting the arguments of Bolton and Harris (1999) to our environment, v must satisfy the Bellman equation

$$0 = \max\{w(p) - v(p); -rv(p) - c + D^2p^2(1-p)^2v''(p)\}, \tag{10}$$

where r is a discount rate that captures the idea that longer judicial processes are penalizing for all parties. The first part of the equation implies that $v(p) \geq w(p)$, which means that the value function always exceeds the welfare obtained by stopping immediately. This is natural, since the option of stopping is available at any time. The second part of the equation describes the evolution of the value function while evidence is accumulated:

$$0 = -rv(p) - c + D^2p^2(1-p)^2v''(p).$$

All solutions to this equation are in closed form when $D^2/r = 3/4$:

$$v(p) = -\frac{c}{r} + \left(A_1 + A_2\left(p - \frac{1}{2}\right)(1-p)^{-2}\right)p^{-\frac{1}{2}}(1-p)^{\frac{3}{2}}, \tag{11}$$

where A_1 and A_2 are free integration constants. For simplicity, in what follows we set $r = 1$ and $D^2 = 3/4$ and vary the cost c.[26]

The region in which evidence is gathered and value functions are determined by the conditions that v is continuous, weakly above w, and when it hits w, it satisfies the smooth pasting property whenever w is continuously differentiable at the hitting point.

Starting with the two-verdict case, one should expect v to coincide with w when p is either close to 0 or close to 1: in this case, there is a high degree of confidence in the defendant's guilt and the value of further evidence gathering is low. Near w's kink (i.e., the threshold p^* at which the sentence switches), however, the value of additional evidence is high, so v should be strictly above w. Thus, it suffices to connect v and w on both sides of p^*. At the connection points, $\hat{p}_1$ and $\hat{p}_2$ such that $\hat{p}_1 < p^* < \hat{p}_2$, v must be equal to w (this is the "value matching" condition) and the derivatives must also coincide (this is the smooth pasting condition).

This imposes four conditions (two value matching and two smooth pasting), and there also four free parameters: the cutoffs $\hat{p}_1$ and $\hat{p}_2$, and the constants A_1 and A_2 arising in equation (11). The result is depicted in Figure 3.

[26]Changing r has an equivalent effect if one changes the signal accuracy parameter D to keep D^2/r constant at $3/4$ and the cost parameter c to keep c/r constant.

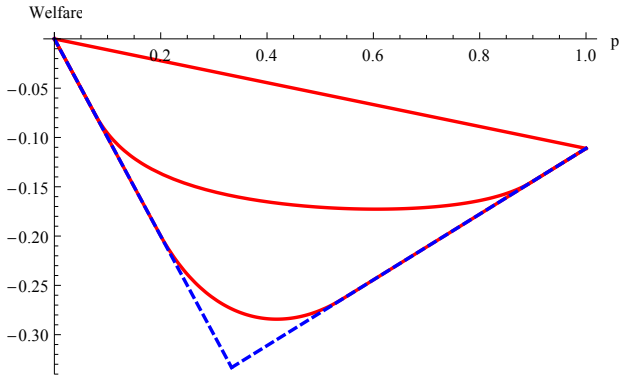

Figure 3: Value function, 2 verdicts, for varying cost levels.

The three-verdict case is more interesting. Around the kink p_2, we still have a two-way smooth connection between w and v, as in the two-verdict case. Around $p_1 = p^*$, however, w is discontinuous, jumping upward from $\underline{w} = -1/3$ to $\bar{w} = -2/9$ as p passes p_1. In this case, if $v(p_1) > \bar{w}$ (the cost is low), then the situation is exactly as in the two-verdict case. Intuitively, the cost is low enough that the intermediate verdict doesn't matter: evidence is gathered until either the not guilty or the guilty verdict is reached. This a situation in which the trial technology is quite accurate, so a two-verdict system suffices.

For larger costs, however, v hits w exactly at $p_1 = p^*$, due to the upward jump. The smooth pasting condition is violated, because the left derivative of v is higher than its right derivative at p_1, and v is equal to w on a right neighborhood of p_1. Intuitively, this kink in the value function reflects the fact that $p_1 = p^*$ was not chosen optimally for the three-verdict system, but rather inherited from the two-verdict system.

The evidence-gathering region now has two parts. When p is below p_1, there is a large incentive to gather evidence, because such evidence can change the sentence from 0 to s_1, and s_1 was tailored to provide a fairer sentence around p_1 than both 0 and s_2. This also implies that not gathering evidence in a right-neighborhood of p_1 is optimal. The second evidence-gathering region is around p_2, as before.[27]

Because the first region violates the smooth pasting condition at p_1, its determination is slightly different. We must determine the threshold $\tilde{p}_0$ at which the region begins, and we know that the region ends at the cutoff p_1. At $\tilde{p}_0$, we have two conditions: the value matching and the smooth pasting conditions. At p_1, however, we only have the value matching condition

[27] As the search cost decreases, the two search regions become connected when $v(p_1) \geq \bar{w}$.

$v(p_1) = \bar{w}$, since the smooth pasting condition is violated. This gives three conditions. There are also three free parameters: the cutoff $\tilde{p}_0$ and the constants $\hat{A}_1$ and $\hat{A}_2$ in (11) for that region. The result is depicted in Figure 4.

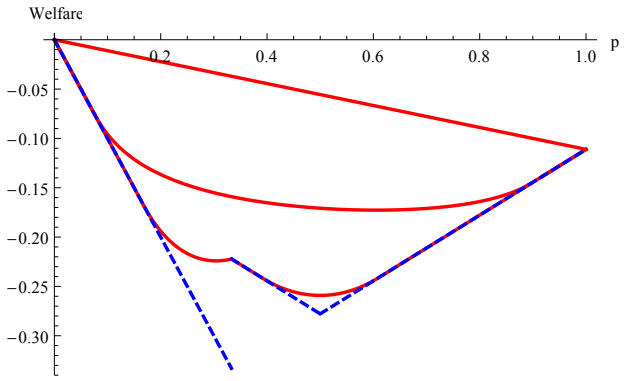

Figure 4: Value function, 3 verdicts, for varying cost levels.

Because the welfare w_3 is always higher than the welfare w_2, it is is straightforward to establish that the value function v_3 in the three-verdict case is (weakly) higher than the two-verdict value function v_2. This matters for high enough cost, i.e., when $v(p_1) = \bar{w}$. In that case, v_3 is strictly above v_2 around p_1, and it is also strictly above v_2 in the second evidence-gathering region, closer to p_2. This implies that the cutoff $\tilde{p}_0$ is lower than the cutoff $\hat{p}_1$ of the two-verdict case, and the right cutoff $\tilde{p}_2$ of the second evidence-gathering region in the three-verdict case is greater than $\hat{p}_2$.

In conclusion, the impact of switching to a three-verdict system by splitting the guilty verdict depends on the evidence gathering cost. When the trial technology is very accurate, the posterior is unlikely to end up in the middle region, so the intermediate verdict has little impact. When finding new evidence is very costly, however, the posterior may end up in the middle region. The third-verdict system then increases the value of gathering evidence in two regions, below p_1 and around p_2, and decreases the value immediately above p_1. Overall, because $\tilde{p}_0 < \hat{p}_1$ and $\tilde{p}_2 > \hat{p}_2$, the three-verdict system results in evidence gathering at more extreme beliefs, where in the two-verdict evidence gathering has already stopped.

5 Intermediate "not guilty" verdict

Suppose now that those defendants who would be acquitted in the current two-verdict system now receive one of two verdicts, which we denote 1 and 2. Both verdicts are associated with no

jail time, i.e., with $s = 0$. Verdict 1, which we refer to as "not guilty," obtains if the posterior is less than some cutoff $p^{iv} < p^*$, and verdict 2, which we refer to as "not proven," obtains if the posterior is between p^{iv} and p^*. We denote by p_i the probability that a defendant is guilty conditional on verdict $i = 1, 2$. A posterior above p^* leads to a conviction and the same sentence s^* as in the two-verdict system.

We assume that society observes the verdict at the end of the trial, but not the posterior regarding the defendant's guilt. The stigmatization associated with being charged and tried is modeled by a cutoff p^s, such that the defendant is stigmatized if the probability he is guilty conditional on the verdict exceeds p^s. We take p^s as exogenous, and assume that convicting a defendant guilty is more demanding than stigmatizing him, so $p^s < p^*$.[28] We also assume that if the defendant is completely cleared in the trial and the public were fully aware of this, then he would not be stigmatized. That is, $\underline{p} < p^s$, where $\underline{p}$ is the lowest possible posterior. An innocent defendant who is stigmatized lowers welfare by $d^i > 0$, and a guilty defendant who is stigmatized increases welfare by $d^g > 0$.[29] We are interested in the optimal cutoff p^{iv} and the conditions under which introducing the additional verdict increases welfare.

The relevant part of the welfare function in the two-verdict system is

$$\lambda \left[W(0, g) + 1_{p^{ng} > p^s} d^g \right] + (1 - \lambda) \left[W(0, i) - 1_{p^{ng} > p^s} d^i \right],$$

where p^{ng} is the probability that a defendant is guilty conditional on being acquitted, since whether an acquitted defendant is stigmatized depends on whether p^s is lower or higher than p^{ng}. We consider these two possibilities below.

Suppose first that $p^{ng} \geq p^s$, so an acquitted defendant in the two-verdict system is stigmatized. For any p^{iv}, it must be that $p_2 \geq p^{ng} \geq p^s$, so the defendant is stigmatized if he is found "not proven" in the three-verdict system. The split can have an effect on social welfare only if $p_1 \leq p^s$, in which case the defendant is not stigmatized if he is found "not guilty" in the three-verdict system. Therefore, consider p^{iv} such that $p_1 < p^s$. Eliminating the stigma when the defendant is found "not guilty" increases the relevant part of the welfare function by

$$-\lambda \sum_{p \leq p^{iv}} f(p|g)\, d^g + (1 - \lambda) \sum_{p \leq p^{iv}} f(p|i)\, d^i.$$

[28] This implies that the analysis of Section 2.1 does not change as a result of the stigma, since a defendant who receives verdicts 1 or 2 is stigmatized.

[29] A similar analysis can be conducted for $d^i \leq 0$ and/or $d^g \leq 0$.

For a given posterior $p \leq p^{iv}$ the increase is

$$-\lambda f\left(p|g\right) d^g + \left(1-\lambda\right) f\left(p|i\right) d^i > 0 \iff \frac{f\left(p|g\right)}{f\left(p|i\right)} < \frac{\left(1-\lambda\right) d^i}{\lambda d^g}. \tag{12}$$

Since $f\left(p|g\right)/f\left(p|i\right)$ increases in the posterior p, a fact we show in Appendix C.1, we obtain the following result.

Proposition 7 *Suppose that being acquitted in the two-verdict system carries a stigma. Then, optimally splitting the acquittal into "not guilty" and "not proven" increases welfare if and only if* $\frac{f(\underline{q}|g)}{f(\underline{q}|i)} < \frac{(1-\lambda)d^i}{\lambda d^g}$.

If the condition in Proposition 7 holds, then the optimal cutoff p^{iv} is the minimum between the highest posterior for which (12) holds and the highest posterior such that $p_1 \leq p^s$. Notice that the condition in Proposition 7 is satisfied more easily if the defendant is more likely to be innocent (λ decreases), the stigma for the innocent increases, or the stigma for the guilty decreases.

Now suppose that $p^{ng} < p^s$, so an acquitted defendant in the two-verdict system is not stigmatized. The split can have an effect on social welfare only if $p_2 > p^s$, in which case the defendant is stigmatized if he is found "not proven" in the three-verdict system. Therefore, consider p^{iv} such that $p_2 > p^s$. Stigmatizing the defendant when he is found "not proven" increases the relevant part of the welfare function by

$$\lambda \sum_{p>p^{iv}} f\left(p|g\right) d^g - \left(1-\lambda\right) \sum_{p>p^{iv}} f\left(p|i\right) d^i.$$

For a given posterior $p > p^{iv}$ the increase is

$$\lambda f\left(p|g\right) d^g - \left(1-\lambda\right) f\left(p|i\right) d^i > 0 \iff \frac{f\left(p|g\right)}{f\left(p|i\right)} > \frac{\left(1-\lambda\right) d^i}{\lambda d^g}. \tag{13}$$

Since $f\left(p|g\right)/f\left(p|i\right)$ increases in the posterior p, we obtain the following result.

Proposition 8 *Suppose that being acquitted in the two-verdict system does not carry a stigma. Then, optimally splitting the acquittal into "not guilty" and "not proven" increases welfare if and only if* $\frac{f(p^*|g)}{f(p^*|i)} > \frac{(1-\lambda)d^i}{\lambda d^g}$.

If the condition in Proposition 8 holds, then the optimal p^{iv} is the maximum between the lowest posterior for which (13) holds and the lowest posterior such that $p_2 \geq p^s$. Notice that the condition in Proposition 8 is satisfied more easily if the defendant is more likely to be guilty (λ increases), the stigma for the innocent decreases, or the stigma for the guilty increases.

6 Reflecting residual doubt in the current justice system

The most explicit inclusion of residual doubt in the U.S. criminal justice system concerns the determination of death sentences. In capital cases, juries must decide, after returning a guilty verdict, whether the defendant should get the death penalty. In this penalty phase, residual or "lingering" doubt may be used as a mitigating circumstance to reject the death penalty.[30] The Capital Jury Project—an academic survey of past jurors in capital cases—has found that lingering doubt was the most important mitigating factor identified by jurors.

There is, however, wide variation in how residual doubt is applied. First, the U.S. penal code (Title 18, §3592) does not explicitly mention residual doubt in its list of mitigating factors, although it does state that mitigating circumstances are not limited to this list. In some cases, jurors are not informed that lingering doubt is a valid mitigating circumstance.[31] In *Franklin v. Lynaugh* (1988), the U.S. Supreme Court rejected a defendant's right to invoke residual doubt at the penalty stage, while in *People v. McDonald* (Supreme Court of Illinois, 1995) a trial judge refused to answer jurors' question on the issue, a decision which was later affirmed by the Supreme Court of Illinois.

Compounding this inconsistency, there is empirical evidence that many jurors get confused with the voting rules used to establish aggravating and mitigating circumstances at the penalty stage. While the unanimity rule is required to find a circumstance aggravating, no such standard exists for mitigating circumstances. The Capital Jury Project found, however, that 45% of jurors failed to understand that they were allowed to consider any mitigating evidence during the sentencing phase of the trial, not just the factors listed in the instructions.[32]

When sentencing is performed by a trial judge, the invocation of residual doubt can be highly controversial. In *State v. Krone* (Arizona Supreme Court, 1995) a trial judge sentenced to life in prison a defendant found guilty of murder, citing doubt about whether he was the true killer.

[30]The more demanding requirement of proving guilty beyond "all doubt" has been discussed in some states, such as the bill proposed in 2003 by then Illinois House Republican leader Tom Cross. Some death penalty advocates have countered that it was impossible to prove anything beyond *all* doubt, and that the bill would in effect rule out the death penalty. Various degrees of lingering doubt have been discussed (e.g., Sand and Rose, 2003) without any mathematical formalism.

[31]See, e.g., *People v. Gonzales and Soliz*, California Supreme Court, 2009.

[32]The CJP's findings concerning jurors' understanding of instructions are summarized at http://www.capitalpunishmentincontext.org/issues/juryinstruct.

In their legal textbook, Dressler and Thomas (2010, pp. 57–61) comment that this decision "borders on the unbelievable." They do not, however, suggest an alternative solution.

In non-capital cases, only five states permit juries to make the sentencing decision. Outside of these states, residual doubt can thus only be expressed by the sentencing judge, whose opinion does not necessarily reflect the views of the jury. Again, residual doubt is not listed as a mitigating factor in sentencing guidelines.

The fact that residual doubt should only be considered in capital cases seems largely arbitrary. Even comparably less serious cases can carry large sentences, resulting in extreme punishments for defendants who are found guilty but for whom residual doubt remains. For example, in *State v. May* (Arizona Superior Court, 2007) a thirty-five-year-old defendant was sentenced to 75 years in jail after being found guilty of touching, in a residential swimming pool, the clothing of four children in the vicinity of their genitals (Nelson, 2013). Jurors had doubts about the guilt of the defendant: they were twice unable to reach a verdict within the first three days of deliberation. The explicit inclusion of residual doubt in sentencing would have likely avoided such an extreme outcome.[33]

Some felonies provide an indirect way of expressing doubt by using the lesser-included-offense rule: juries can return a manslaughter verdict, rather than a first- or second-degree murder verdict, or a larceny verdict instead of a robbery verdict. However, each of these verdicts corresponds to a precise charge (e.g., whether premeditation and malice aforethought were involved) and doubt about a particular charge can only be imperfectly expressed by returning a guilty verdict on a lower count. These instruments only offer a limited and, in fact, improper, way of reflecting residual doubt. Furthermore, the less-included-offense rule is not a constitutional right of the defendant; its application is therefore to some extent arbitrary and depends on the inclination of the jury (see Mascolo (1986)).

Even when the lesser-included-offense rule does not apply, residual doubt may be reflected by returning a guilty verdict only on a subset of the charges brought against the defendant. There is anecdotal evidence that such compromise is sometimes used by the jurors to reflect doubt. In

[33]Capital sentences are unique in their irreversibility, which creates an additional reason for avoid this sentence in case of lingering doubt: exonerating evidence may come after the execution of defendant, preventing any release and compensation. In practice, this fundamental difference is attenuated by the fact that death-row defendants spend many years in jail before their execution until all recourses have been exhausted, while non-capital defendants serving long sentences may die in jail, which also prevents any release or compensation.

the aforementioned State v. May, for instance, Nelson (2013) notes that "it seems likely that the defendant molested either all of the children or none of them. So why did the jury ultimately reach a verdict of guilty on five counts and not guilty on two? The answer is that the jurors compromised." Dropping some charges is, however, a very coarse instrument to incorporate residual doubt: for example, this approach cannot be used to reduce the sentence of a defendant facing a single but severe count, while it may be used for a defendant facing several counts, the sum of which adds to the same aggregate maximal sentence as in the single-count case.[34] Even when it is feasible, the approach exposes the defendant to another idiosyncratic component of the jury—whether it is sophisticated or willing enough to use this compromise strategy—introducing a source of jury heterogeneity in trial outcomes even for otherwise identical cases.[35]

The U.S. justice system incorporates residual doubt about a defendant's guilt in two other ways. First, a defendant found not guilty in a criminal trial may still be found guilty in a civil suit, which is uses the more permissive preponderance-of-evidence standard of proof. However, civil suit sentences carry no jail time and thus may be more limited in preventing recidivism. Furthermore, the connection between criminal and civil trials is generally limited, preventing any coordination and coherent decision across these trials. Second, residual doubt variations also imply different likelihoods of post-trial events such as successful appeals and exonerations, which affect the defendant's ultimate punishment. These events are largely beyond the control of the first court and are not a close substitute for the additional verdicts introduced here.

In summary, the current criminal justice system includes various ways of reflecting residual doubt in outcomes and it appears that these ways are used purposefully by some actors of the system. However, these ways are largely arbitrary, inconvenient, and uncoordinated. This paper proposes a structured, systematic approach for the consideration of residual doubt in criminal justice decisions and explicit designs which are shown to improve welfare in many settings.

[34]The set of charges leveled at the defendant may also be affected by the strategic decisions of the prosecutor, which increases the prosecutor's power and adds to the complexity of this problem.

[35]It should also be noted that under the current law, such compromise is actually illegal if it results from a bargaining between pro-acquittal and pro-conviction jurors. Such an arrangement currently violates the rights of the defendant if the pro-acquittal jurors still believe that the defendant should be found not guilty (Mascolo (1986)).

7 Implementation and jurors' reactions to additional verdicts

Implementation: verdicts vs. sentences

Formalizing the intermediate sentence introduced in this paper as an intermediate *verdict* is consistent with the not-proven verdict, discussed in Section 5, used by some criminal justice systems. In this formulation, the jury must decide, according to some collective rule, among the three verdicts.

An alternative "two-step" implementation maintains the current separation between the fact-finding and sentencing stages. The verdict outcome is still binary ("guilty" or "not guilty"), and residual doubt is expressed in the form of intermediate sentences decided in the sentencing stage.

The second implementation presents a significant advantage: in principle, the jury can be given exactly the same instructions as in the current system, which allows to cleanly split the set of cases which would receive a "guilty" verdict under the current system into multiple sentence levels reflecting the strength of evidence, and thus leaves unchanged the probability of acquitting the defendant.

Intermediate sentences can be decided in a variety of ways, which may involve a sentencing judge, sentencing guidelines (e.g., automatically rule out the death penalty if the evidence is solely based on a confession), or a jury.

Regardless of the implementation, a potential concern is how the jury may react to additional verdicts. The remainder of our discussion focuses on this issue.

Jurors' reaction to additional verdicts

Jury decisions involve collective and psychological considerations: jurors may have limited and uneven ability to understand jury instructions or interpret the evidence, have varied tolerance for erroneous convictions and acquittals, and are subject to individual biases and to persuasion and group-think dynamics, to cite only a few issues. Even abstracting from these issues, jury decisions are difficult to analyze.[36]

[36]Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1996, 1997), and Gerardi and Yariv (2007) identify important informational effects, which may arise even when all jurors have identical preferences. A central mechanism in this literature is that, conditional on being pivotal in a vote, a rational juror may put so

The literature on criminal trial design varies from fully rational to completely reduced-form models of jury behavior. At the most "rational" extreme, Lee (2015) considers jurors who perfectly take into account how prosecutors select the pool of defendants who go to trial. Prosecutors can influence this pool by choosing the plea sentence that they propose to defendants before the trial.[37] Other papers on trial design (Kaplow (2011), Daughety and Reinganum (2015a,b), Da Silveira (2015), Silva (2015)) abstract from any jury decision, focusing on reduced-form thresholds or on a mechanism design approach without jurors.

A key observation is that our Propositions 1 and 2 continue to hold under the two-step implementation mentioned above, provided that jurors are given the same instructions as in the current system to decide between the guilty and not-guilty verdicts, and react to these instructions in the same way, no matter how imperfect, as they currently do. No matter how "tough on crime" or otherwise biased each juror is, what voting, persuasion or other collective processes are at play, all these components would play out in exactly the same way at the fact-finding stage, under a standard binary verdict, as in the first step of the two-step approach, guaranteeing that no more defendants are found guilty in the three-verdict system than in the current one.

The main question, therefore, is to what extent jurors would know and incorporate in the fact-finding stage the fact that residual doubt may be used as a mitigating factor in the sentencing stage.

In practice, there is little evidence that jurors incorporate sentencing considerations into their verdict decisions. On the contrary, in recent history judicial practice has been to keep the jury uninformed about the the punishment faced by the defendant (Sauer (1995)). In *United States v. Patrick* (D.C. Circuit, 1974), the court affirmed that the jury's role is limited to a determination of guilt or innocence. Instructions entirely focus on describing the procedure for finding facts. In many cases—such as *People v. May* above—jurors are unaware of the minimum-punishment guidelines relevant for the case.

There is also empirical evidence that harsher sentences do not result in lower conviction rates. In a study of non-homicide violent case-level data of North Carolina Superior Courts, Da Silveira (2015) finds that the probability of conviction of defendants going to trial in fact

much weight on other jurors' signals that he significantly discounts, and potentially discards, his own information.

[37]The approach presumes that jurors are aware of the plea sentence offered to the defendant. In practice, the jury is often instructed to consider only the evidence produced at trial.

increases with the sentence that they face.[38] Such a correlation cannot be easily explained away by prosecutor behavior: if, in particular, prosecutors attached more importance to obtaining a conviction when the case is more severe, they would send to trial defendants who are more likely to be found guilty and obtain a guilty plea from the other ones, and one would expect the probability of plea settlements to increase with the severity of the trial sentence. This relation seems contradicted by the data.[39]

More generally, there is strong evidence that jurors have a limited understanding of the sentences faced by defendants. For example, the aforementioned Capital Jury Project found that most jurors "grossly underestimated" the amount of time spent in jail entailed by a guilty verdict. It is reasonable to believe that jurors would be as unaware of, say, maximum-sentencing guidelines, as they currently are of minimum-sentencing guidelines.

Finally, if contrary to expectations jurors incorporated the intermediate verdict into their decision, they might adopt a lower standard of proof to convict defendants, knowing that the corresponding cases would result in a lower sentence than in the current system. To the extent that jurors did so with the social welfare objective in mind, such a change would likely be beneficial—indeed, Proposition 4 shows that the optimal three-verdict system has this feature. Jurors may, however, have their own objective in mind. For example, they may worry about the length of deliberation, and be willing to continue deliberation only if the social value of doing so is high. The analysis of Section 4 suggests that this value is not lowered by the introduction of a third verdict, and may in fact be higher for a wider range of beliefs.

[38]Da Silveira's analysis excludes the most and least severe cases to focus on a relatively homogeneous pool of cases.

[39]Elder (1989) finds evidence that circumstances that may aggravate punishment *reduce* the probability of settlement. Similarly, Boylan (2012) finds that a 10-month increase in prison sentences raises trial rates by 1 percent.

References

Athey, S. (2002) "Monotone Comparative Statics under Uncertainty," *Quarterly Journal of Economics*, Vol. 117, pp. 187–223.

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

Becker, G. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy,* Vol. 76, pp. 169–217.

Bolton, P., Harris, C. (1999) "Strategic Experimentation," *Econometrica*, Vol. 67, pp. 349–374.

Boylan, R. (2012) "The Effect of Punishment Severity on Plea Bargaining," *Journal of Law and Economics*, Vol. 55, pp. 565–591.

Bray, S. (2005) "Not Proven: Introducing a Third Verdict," *University of Chicago Law Review*, Vol. 72, pp. 1299–1329.

Burns, R. (2009) *The Death of the American Trial*, University of Chicago Press.

Da Silveira, B. (2015) Bargaining with Asymmetric Information: An Empirical Study of Plea Negotiations," *Working Paper*, Washington University.

Daughety, A., Reinganum, J. (2015a) "Informal Sanctions on Prosecutors and Defendants and the Disposition of Criminal Cases," *Working Paper*, Vanderbilt University.

Daughety, A., Reinganum, J. (2015b) "Selecting Among Acquitted Defendants: Procedural Choice vs. Selective Compensation," *Working Paper*, Vanderbilt University.

Dressler, J., Thomas, G. (2010) "Does the Process Protect the Innocent," in *Criminal Procedure: Prosecuting Crime*, Fourth Edition, West Academic Publishing.

Elder, H. (1989) "Trials and Settlement in the Criminal Courts: an Empirical Analysis of Dispositions and Sentencing," *Journal of Legal Studies*, Vol. 18, pp. 191–208.

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

Feddersen, T., Pesendorfer, W. (1997) "Voting Behavior and Information Aggregation in Elections with Private Information," *Econometrica,* Vol. 65, pp. 1029–1058.

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

Grogger, J. (1992) "Arrests, Persistent Youth Joblessness, and Black-White Employment Differentials," *Review of Economics and Statistics*, Vol. 74, pp. 100–106.

Grogger, J. (1995) "The Effect of Arrest on the Employment and Earnings of Young Men," *Quarterly Journal of Economics*, Vol. 90, pp. 51–72.

Gross, S., O'Brien, B., Hu, C., and E. Kennedy (2014) "Rate of False Conviction of Criminal Defendants who are Sentenced to Death," *Proceedings of the National Academy of Sciences*, Vol. 111, pp. 7230–7235.

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

Kaplow, L. (2011) "On the Optimal Burden of Proof," *Journal of Political Economy*, Vol. 119, pp. 1104–1140.

Lee, S. (2014) "Plea Bargaining: On the Selection of Jury Trials," *Economic Theory*, Vol. 57, pp. 59–88.

Lott, J. (1990) "The Effect of Conviction on the Legitimate Income of Criminals," *Economics Letters*, Vol. 34, pp. 381–385.

Mascolo, E. (1985) "Procedural Due Process and the Lesser-Included Offense Doctrine," *Albany Law Review*, Vol. 50, pp. 263–304.

Milgrom, P., Segal, I. (2002) "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, Vol. 70, pp. 583–601.

Milgrom, P., Shannon, C. (1994) "Monotone Comparative Statics," *Econometrica*, Vol. 62, pp. 157–180.

Nelson, W. (2013) "Political Decision Making by Informed Juries." *William and Mary Law Review*, Vol. 55, pp. 1149–1166.

Quah, J., and Strulovici, B. (2012) "Discounting, Values, and Decisions," *Journal of Political Economy*, Vol. 121, pp. 898–939.

Sand, L., Rose, D. (2003) "Proof Beyond All Possible Doubt: Is there a Need for Higher Burden of Proof When the Sentence May Be Death," *Chicago-Kent Law Review*, Vol. 78, pp. 1359–1376.

Sauer, K. (1995) "Informed Conviction: Instructing the Jury About Mandatory Sentencing Consequences," *Columbia Law Review,* Vol. 95, pp. 1232–1272.

Silva, F. (2015) "The Optimal Design of a Criminal Justice System," *Working Paper*, University of Pennsylvania.

Stigler, G. (1970) "The Optimum Enforcement of Laws," *Journal of Political Economy,* Vol 78, pp. 526–536.

A Proof of Proposition 4

Let (p^*, s^*) and $(p_1^*, p_2^*, s_1^*, s_2^*)$ respectively denote the optimal parameters for the two- and three-verdict systems. We recall our standing assumptions that the conditional welfare functions $W(s, g)$ and $W(s, i)$ are concave in s and the conditional posterior distributions $F(p|g)$ and $F(p|i)$ are absolutely continuous. Finally, we assume that the functions $W(s, g)$ and $W(s, i)$ are twice differentiable in s to warrant the application, below, of the Implicit Function Theorem.

A.1 Comparison of p^* and p_1^*

Proposition 9 *The optimal two-verdict system results in more acquittals than the optimal three-verdict system:* $p^* \geq p_1^*$.

Proof. Let $p_1^*(p_2)$, $s_1^*(p_2)$, $s_2^*(p_2)$ denote the optimal parameters of a three-verdict system in which the higher cut-off p_2 is given exogenously. When $p_2 = 1$, the solution corresponds to the optimal two-verdict system, since the domain $[p_2, 1]$ over which the sentence s_2 is applied collapses. The proposition thus follows if we can show that $p_1^*(p_2)$ is nondecreasing in p_2. To show this result, notice that the choice of s_2 has no effect on the optimal choice of p_1, so that the part of the welfare objective corresponding to s_2 can be dropped from the analysis. The resulting objective function is

$$\lambda\left[(F(p_2|g) - F(p_1|g))W(s_1, g) + F(p_1|g))W(0, g)\right] + (1-\lambda)\left[(F(p_1|g) - F(p_1|i))W(s_1, i) + F(p_1|i)W(0, i)\right]. \quad (14)$$

Let $\mathcal{W}_{reduced}(p_1, p_2)$ denote the value of this objective optimized with respect to s_1. To show that $p_1^*(p_2)$ is nondecreasing, it is enough to show that $\mathcal{W}_{reduced}$ is supermodular in (p_1, p_2) or, equivalently, that its cross-partial derivative is everywhere nonnegative.[40] Applying the Envelope Theorem (see, e.g., Milgrom and Segal (2002)), we have

$$\frac{\partial^2 \mathcal{W}_{reduced}}{\partial p_1 \partial p_2} = \frac{\partial s_1^*(p_1, p_2)}{\partial p_1}\left[\lambda f(p_2|g)W'(s_1^*(p_1, p_2), g) + (1-\lambda)f(p_2|i)W'(s_1^*(p_1, p_2), i)\right] \quad (15)$$

where $s_1^*(p_1, p_2)$ denote the sentence which maximizes (14) given any values of p_1 and p_2.

We begin by showing that the first factor of (15), $\frac{\partial s_1^*(p_1,p_2)}{\partial p_1}$, is positive. Since $s_1^*(p_1, p_2)$ satisfies the first-order condition[41]

$$\lambda\left[(F(p_2|g) - F(p_1|g))W'(s_1^*(p_1, p_2), g)\right] + (1-\lambda)\left[(F(p_1|g) - F(p_1|i))W'(s_1^*(p_1, p_2), i)\right] = 0, \quad (16)$$

the Implicit Function Theorem implies that $\frac{\partial s_1^*(p_1,p_2)}{\partial p_1}$ has the same sign as

$$-\lambda f(p_1|g)W'(s_1^*(p_1, p_2), g) - (1-\lambda)f(p_1|i)W'(s_1^*(p_1, p_2), i).$$

Comparing this expression with (16), the claim follows $\frac{f(p_1|g)}{f(p_1|i)} < \frac{F(p_2|g)-F(p_1|g)}{F(p_2|i)-F(p_1|i)}$ (by the Monotone Likelihood Ratio Property (MLRP) established by Proposition (13)) and the fact that $W'(s_1^*(p_1, p_2), g) > 0 > W'(s_1^*(p_1, p_2), i)$.[42]

We now show that the second factor of (15), $\lambda f(p_2|g)W'(s_1^*(p_1, p_2), g) + (1-\lambda)f(p_2|i)W'(s_1^*(p_1, p_2), i)$, is positive. This follows from (16), the fact—implied by the MLRP—that $\frac{f(p_2|g)}{f(p_2|i)} > \frac{F(p_2|g)-F(p_1|g)}{F(p_2|i)-F(p_1|i)}$, and the inequalities $W'(s_1^*(p_1, p_2), g) > 0 > W'(s_1^*(p_1, p_2), i)$. ■

[40] See Milgrom and Shannon (1994).

[41] As in Section 2.1, $W'(s, g)$ and $W'(s, i)$ denote the derivatives of $W(s, g)$ and $W(s, i)$ with respect to s.

[42] These strict inequalities are implied by the monotonicity of the welfare functions over the relevant interval, combined with their strict concavity.

A.2 Comparison of p^* and p_2^*

Proposition 10 *The optimal two-verdict system convicts more often than the optimal three-verdict system gives the higher sentence: $p^* \leq p_2^*$.*

Proof. Fix $p_1 = p_1^*$ and $s_1 = s_1^*$. We have $p_1 \leq p^*$, from Proposition 9. We need to show that the optimal cutoff $p_2(p_1, s_1)$ in the three-verdict system, taking p_1 and s_1 fixed at these values, is greater than p^*. If $s_1 = 0$, the three-verdict system reduces to a two-verdict system with cutoff p_2. Hence, the optimal cutoff is $p_2 = p^*$ and the claim holds. Assume now that $s_1 > 0$. If $p_1 = p^*$, the claim also holds trivially since p_2 must lie above p_1^*. Suppose, therefore, that $p_1 < p^*$ and assume by way of contradiction that $p_2^* < p^*$. Consider any p_2 lying in (p_1, p^*) and let

$$\hat{W}(p_2) = \lambda \{F([p_1, p_2]|g)W(s_1, g) + F([p_2, 1]|g)W(s_2(p_2), g)\} + (1-\lambda) \{F([p_1, p_2]|i)W(s_1, i) + F([p_2, 1]|i)W(s_2(p_2), i)\}$$

which is the part of the social welfare function that involves p_2, where the notation $s_2(p_2)$ reflects the fact the optimal sentence over the interval $[p_2, 1]$ depends only on the threshold p_2, not on s_1 or p_1. We will show that $\hat{W}(p^*) \geq \hat{W}(p_2)$ for all $p_2 \in (p_1, p^*)$, which will establish that the welfare-maximizing threshold is greater than p^* and yield the desired contradiction.[43]

To show this, fix any $p_2 \in (p_1^*, p^*)$. The difference $\hat{W}(p^*) - \hat{W}(p_2)$ can be decomposed as follows: when the posterior lands anywhere between p_1 and p_2, the systems that achieve welfare levels $\hat{W}(p^*)$ and $\hat{W}(p_2)$ both assign the sentence s_1 to the defendant. This part of the welfare thus cancels out from the difference. Above p_2, $\hat{W}(p_2)$ is computed using the optimal single sentence, $s_2(p_2)$, between p_2 and 1, while $\hat{W}(p^*)$ is computed using the optimal single sentence, $s_2(p^*)$, between p^* and 1, and using s_1 over the interval $[p_2, p^*]$. When $s_1 = 0$, we are back to comparing the cutoffs p^* and p_2 in the two-verdict system and, by optimality of p^* for two-verdict systems, the difference is positive. When s_1 is increased, the only change is that over $[p_2, p^*]$, $\hat{W}(p^*)$ is now computed using a positive sentence rather than a 0 sentence. Clearly, this change is welfare-improving as long as s_1 is not too high. There only remains to show that for $s_1 = s_1^*$, the change indeed improves welfare.

To achieve this, we start with the following observation. For any $p < p'$, let

$$W(s; [p, p']) = \lambda F([p, p']|g)W(s, g) + (1 - \lambda)F([p, p']|i)W(s, i) \tag{17}$$

denote the part of the welfare function that concerns the posterior lying in $[p, p']$ when the sentence is s. Since $W(s, g)$ and $W(s, i)$ are both concave in s, so is $W(s; [p, p'])$. This implies that 17 decreases as the sentence moves away (in either direction) from the optimal sentence $s^*([p, p'])$.

Consider the optimal sentence $\hat{s} = s(p_2, p^*)$ over our interval of interest, $[p_2, p^*]$. If we show that $s_1^* \leq \hat{s}$, the previous observation will imply that setting the sentence to s_1^* over the interval $[p_2, p^*]$ is indeed better than setting it to zero, which will conclude the proof.[44] The proof that $s_1^* \leq \hat{s}$ is based on the following lemma:

Lemma 1 *The sentence $s^*(p_1, p_2)$ which maximizes the objective*

$$\lambda F([p_1, p_2]|g)W(s, g) + (1 - \lambda)F([p_1, p_2]|i)W(s, i) \tag{18}$$

is increasing in p_1 and p_2.

[43] In principle, there could exist multiple optimal thresholds. The argument presented here implies that there must exist at least one optimal threshold that lies above p^*.

[44] Since the posterior distributions are continuous, the value of the sentence at interval extremities is unimportant. We abuse notation slightly by using closed intervals everywhere.

Proof. The MLRP property implies that the ratio $\beta(p_1, p_2) = F([p_1, p_2]|g)/F(p_1, p_2]|i)$ is increasing in p_1 and p_2 over the set $\{p_1, p_2) \in [0,1]^2 : p_1 < p_2\}$. Dividing the objective (18) by $F([p_1, p_2]|i)$ yields $\lambda\beta(p_1, p_2)W(s, g) + (1-\lambda)W(s, i)$. This modified objective satisfies the single-crossing property in $(s; p_1)$ because β is increasing in p_1 and positive while $W(s, g)$ is increasing in s.[45] This implies that $s^*(p_1, p_2)$ is increasing in p_1. By the same reasoning, it is also increasing in p_2. ■

To conclude the proof of Proposition 10, recall our contradiction hypothesis that $p_2^* \leq p^*$. Since i) $s_1^* = s^*(p_1^*, p_2^*)$, ii) $\hat{s} = s^*(p_2, p^*)$, iii) $p_1^* \leq p_2$ and $p_2^* \leq p^*$, Lemma 1 implies that $s_1^* \leq \hat{s}$, which shows the desired inequality. ■

A.3 Comparison of s^*, s_1^* and s_2^*

The ordering of the optimal sentences then follows immediately from Lemma 1 and the previous two propositions. Intuitively, the optimal sentence reflects how likely the agent is guilty. So 'higher' sets of priors will lead to a longer sentence.

B Parameters for the welfare functions of Section 4

We set to 1 the ideal sentence $\bar{s}$ for the guilty and use quadratic loss functions: $W(s|g) = -(1-s)^2$, $W(s|i) = -s^2$. We also assume that the prior is equal to 1/2: the defendant is equally likely to be guilty or innocent ex ante. To obtain simple expressions for the optimal cutoffs and sentences, we reverse-engineer the signal structure. Recall that the optimal cutoff is given by the indifference condition

$$p^*W(s^*, g) + (1-p^*)W(s^*, i) = p^*W(0, g),$$

or $p^*(1-(s^*)^2) + (1-p^*)(-(s^*)^2) = p^*$. The optimal sentence is given by the first-order condition deriving from

$$s^* \in \arg\max_s \frac{1}{2}Pr(p \geq p^*|g)W(s|g) + \frac{1}{2}Pr(p \geq p^*|i)W(s|i),$$

i.e., $(1-F(p^*|g))(1-s^*) = (1-F(p^*|i))s^*$. By choosing $F(\cdot, g)$ and $F(\cdot, i)$ so that the ratio $q = \frac{1-F(p|i)}{1-F(p|g)}$ is equal to 1/2 when evaluated at $p = 1/3$, we verify that $p = 1/3$ and $s = 2/3$ solve the problem. Note that q must be less than 1, from MLRP.

With three verdicts, we impose the restrictions $p_1 = 1/3$ and $s_2 = 2/3$— so that we are indeed splitting the guilty verdict, and not increasing the guilty sentence, and optimize over the remaining two parameters, p_2 and s_1. These parameters are again characterized by the indifference equation for p_2, given the sentences s_1 and s_2 that are given above and below p_2,

$$p_2W(s_1, g) + (1-p_2)W(s_1, i) = p_2W(s_2, g) + (1-p_2)W(s_2, i),$$

and by the optimality condition for s_1, which is

$$s_1 \in \arg\max_s \frac{1}{2}Pr(p \in [p_1, p_2]|g)W(s|g) + \frac{1}{2}Pr(p \in [p_1, p_2]|i)W(s|i),$$

which yields the first-order condition

$$F([p_1, p_2]|g)(1-s_1) = F([p_1, p_2]|i)s_1.$$

[45] In fact, taking the derivative with respect to s yields an increasing function of p_1, showing that supermodularity in (s, p_1), which implies the single-crossing property. See, e.g., Milgrom and Shannon (1994).

Again doing reverse engineering, we choose $F(\cdot|g)$ and $F(\cdot,i)$ so that the ratio $q' = \frac{F([p_1,p_2]|i)}{F([p_1,p_2]|g)}$ evaluated at $p_1 = 1/3$ and $p_2 = 1/2$ be equal to 2. With this condition, $s_1 = 1/3$ and $p_2 = 1/2$ satisfy all conditions. Note that the ratio q' must be greater than q, by MLRP.

This yields the welfare functions $w_2(p) = w_3(p) = -p$ for $p < 1/3$, $w_2(p) = -p/9 - (1-p) \times 4/9$ for $p \geq 1/3$, and $w_3(p) = -p/9 - (1-p) \times 4/9$ for $p \geq 1/2$, and $w_3(p) = -p\frac{4}{9} - (1-p)\frac{1}{9}$ for $p \in [1/3, 1/2)$.

C Foundation of the Bayesian Conviction Model

We now study whether actual court proceedings can be translated into a Bayesian updating process and a threshold. We address this by considering an evidence-based trial technology. There is a set X of evidence elements, and "evidence collection" refers to a subset of X. The court technology is a mapping $D : 2^X \to \{G, N\}$, which for every evidence collection decides whether the defendant is guilty or not guilty.[46] Distributions P_θ on 2^X, for $\theta \in \{g, i\}$, describe the probability that different evidence collections arise conditional on the defendant being actually guilty or innocent. We assume that both distributions have full support. Letting π_θ^k denote the probability that a defendant of type θ receive verdict k, we have $\pi_\theta^k = P_\theta\left(D^{-1}(k)\right)$ for each type θ and verdict k in $\{G, N\}$. Recall that $\pi_i^G < \pi_g^G$, i.e., $P_i\left(D^{-1}(G)\right) < P_g\left(D^{-1}(G)\right)$, and that λ is the prior that the defendant is guilty. We ask several questions.

1. Given D, P_i, P_g, and λ, can D be rationalized as the result of Bayesian updating with a threshold on the posterior for determining guilt? At a minimum, this would require D to respect "incriminating" and "exculpatory" evidence sets, which are determined by whether they indicate that the defendant is more likely to be guilty than innocent.

2. Given D and λ, can P_i and P_g be chosen to rationalize D as the result of Bayesian updating with a threshold on the posterior for determining guilt?

3. Given λ, can D, P_i, and P_g be chosen to rationalize D as the result of Bayesian updating with a threshold on the posterior for determining guilt?

To answer these questions, we formally order defendant types i and g so that $i < g$, and we order verdicts as $N < G$. Then, we say that D ***can be rationalized*** as the result of Bayesian updating with a threshold on the posterior if for every $E, E' \subseteq X$ we have $D(E) < D(E')$ if and only if the posterior that the defendant is guilty is higher under E' than under E, i.e.,

$$\frac{\lambda P_g(E)}{\lambda P_g(E) + (1-\lambda) P_i(E)} < \frac{\lambda P_g(E')}{\lambda P_g(E') + (1-\lambda) P_i(E')}.$$

This condition is equivalent to $\lambda P_g(E)\left(\lambda P_g(E') + (1-\lambda) P_i(E')\right) < \lambda P_g(E')\left(\lambda P_g(E) + (1-\lambda) P_i(E)\right)$ and, after rearranging, to

$$\frac{P_g(E)}{P_i(E)} < \frac{P_g(E')}{P_i(E')}.$$

The likelihood ratios are thus ordered independently of λ. For every evidence set $E \subseteq X$, denote by $r(E) = P_g(E)/P_i(E)$ its likelihood ratio. This shows the following proposition.

Proposition 11 *D can be rationalized if and only if for every $E, E' \subseteq X$ the following holds:*

$$r(E) \leq r(E') \Rightarrow D(E) \leq D(E').$$

[46] The analysis can be generalized to stochastic decisions.

While we started with a Bayesian definition of rationalizability, this concept is in fact non-Bayesian: it is purely based on the likelihood ratio of guilty given the observed evidence and, in particular, is independent of any prior.

Equipped with this result, we can answer the questions above. For 1, the answer is "yes" if and only if

$$\max\{r(E) : D(E) = N\} < \max\{r(E) : D(E) = G\}. \tag{19}$$

For 2, the answer is "yes:" choose P_g and P_i so that (19) holds. Since 2 implies 3, that answer to 3 is also "yes."

Incriminating and exculpatory evidence: definitions and properties

When D can be rationalized, we say that evidence $e \in X$ is **D-incriminating** if for every $E \subseteq X$ with $e \notin E$, $D(E) = g$ implies that $D(E \cup \{e\}) = g$. We say that evidence $e \in X$ is *P-incriminating* if for every $E \subseteq X$ with $e \notin E$ we have that $r(E) \leq r(E \cup \{e\})$. Decision- and belief-based notions of exculpatory evidence are defined similarly. The following result.

Proposition 12 *If D is rationalized by P, any P-incriminating evidence is also D -incriminating.*

The reverse need not hold: one can easily construct examples in which some evidence collection E suffices to convict the defendant (i.e., $D(E) = g$) and the additional piece of evidence e reduces the 'guilt' ratio ($r(E \cup \{e\}) < r(E)$), but not enough the change the decision ($D(E \cup \{e\}) = g$).

Our definition and characterization of rationalization extend without change to probabilistic functions D, in which the image of D is the probability that the defendant is found guilty.

C.1 Ordering posterior distributions with the MLRP

In the Bayesian conviction model, the posterior belief is formed by combining a prior with the signals observed about the defendant. One may view each evidence collection E as a signal, and signals may be ordered according to the likelihood ratio $r(E)$. The distributions P_i and P_g over evidence collections can then be mapped into distributions over likelihood ratios r. In a Bayesian conviction model, only the likelihood ratio matters for the decision, and one can thus without loss identify any signal with r. Thus, without loss, signals may be ranked according to this likelihood ratio. Let R_g and R_i denote the distributions of r, conditional on being guilty and innocent, respectively. When the signal distributions, conditional on being guilty or innocent, are continuous, let ρ_g and ρ_i denote their densities. By construction, we have $\rho_g(r)/\rho_i(r) = r$. In statistical terms, this means that R_g and R_i are ranked according the MLRP: the ratio of their density is increasing in the signal. Moreover, because the posterior $p(r)$, given a signal r, is equal to the conditional probability of $\theta = g$ given r, it inherits the MLRP.[47] Let F_g and F_i denote the distributions of p, conditional on being guilty and innocent, respectively, and let f_g and f_i denote the densities of F_g and F_i (which exist as long as R_g and R_i are continuous), we have $f_g(p)/f_i(p)$ is increasing in p.

Proposition 13 *Suppose that both signal distributions, conditional on being guilty and innocent, are continuous. Then both distributions of the posterior p are continuous, and their density functions satisfy the MLRP.*

This property, which holds without loss (except for the continuity assumption, of a technical nature), plays a key role for Lemma 1 and the subsequent results.

[47] This fact is well-known and straightforward to establish.: if θ is the state of the world, r is a signal, and the conditional distributions $\rho(r|\theta)$ are ranked according to MLRP, then the posterior distributions $\rho(\theta|r)$ are also ranked according to the MLRP.

D The suboptimality of plea bargains with excessive trial sentences

We introduce a model in which some innocent defendants indeed take the plea. Following GK, we achieve this by introducing two types of innocent defendants, which vary according to their degree of risk aversion. To simplify the analysis, we assume that there are three types of defendants in equal proportion: risk neutral guilty defendants with utility $u(s) = -s$, risk neutral innocent defendants with the same utility, and risk averse innocent defendants with a piecewise linear utility function given by $u(s) = -\frac{3}{16}s$ for $s \leq 16$ and $u(s) = -3 - 2(s - 16)$ for $s \in [16, 20]$. Again for simplicity, we assume that the social welfare as a function of the guilty defendant's punishment is linear with a peak at 20 years: $W(s, g) = -|s - 20|$. We thus only consider sentences lower than the sentence $\bar{s} = 20$ that is optimal if the defendant is known to be guilty.

Finally we suppose that the trial can generate two types of evidence against the defendant, weak or strong. A guilty defendant generates strong evidence with probability 30% and weak evidence with probability 50%. An innocent defendant generates (regardless of his risk aversion) strong evidence with probability 10% and weak evidence with probability 30%. When no evidence is found against the defendant, he is acquitted.

We now show that plea bargaining with two verdicts when the guilty sentence is excessively high is worse than a three-verdict system as in Section 2.1 that keeps the excessively high sentence for the verdict associated with strong evidence.

Because of the linear structure of payoffs, it is easy to show that the only relevant sentence levels are $s_1 = 16$ and $s_2 = 20$. The following facts are easy to establish in this example:

- In a two-verdict system without a plea, it is optimal to punish the defendant for either type of evidence (weak or strong), and the optimal sentence is $s_1 = 16$;
- The same is true in an optimal two-verdict system with a plea, and only the guilty defendant takes the plea;
- If, however, the conviction sentence is suboptimally set to $s_2 = 20$ at the trial stage (which is the ex post optimum if the defendant is indeed guilty), then the optimal plea is $s^b = 0.8 * s_2 = 16$, and both guilty and the risk averse innocent defendants take the plea.
- Subject to keeping a high sentence equal to $s_2 = 20$, the three-verdict system that gives a sentence of $s_1 = 16$ if weak evidence is presented, and $s_2 = 20$ if strong evidence is presented is optimal and yields a higher expected welfare than the two-verdict system with a plea that has a trial conviction sentence of $s_2 = 20$.

This result shows that the introduction of an intermediate verdict with a lower sentence may be more efficient than a plea to counteract the effects of a suboptimally high sentence for the guilty. This illustrates how ethical considerations (here, providing the right ex post punishment if the defendant is guilty) shape the optimal verdict system: in a purely utilitarian world, a suboptimally high guilty sentence would be reduced (here, to 16) and plea bargains may be optimal. If, however, it is difficult to reduce the guilty sentence, due to political or other considerations, plea bargaining not be the best solution.

Another reason plea bargains may be suboptimal is that an innocent defendant may think that his likelihood of being convicted is higher than it really is. Revisiting the example, suppose that the risk averse innocent defendant erroneously believes that the probability of weak evidence being found against him is 75%. Then he may prefer to take the plea rather than run the risk of being found guilty in trial. In this case, even if the guilty sentence is set to $s = 16$, welfare is suboptimal compared to a three verdict system.