European Journal of Applied Mathematics (2026), 1–22
doi:10.1017/S0956792525100296

PAPERS

Modelling crime response to deterrence: Existence of solutions, optimal policies, and fairness

Yosia Nurhan and Martin B. Short

School of Mathematics, Georgia Institute of Technology, Atlanta, GA, 30332, USA
Corresponding author: Martin Short; Email: mbshort@math.gatech.edu

Received: 22 May 2025; **Revised:** 08 December 2025; **Accepted:** 08 December 2025

Keywords: crime modelling; game theory; algorithmic fairness

2020 Mathematics Subject Classification: 91-10, 91B74 (Primary); 91D10, 91B14 (Secondary)

Abstract
We study a model in which rational agents decide whether or not to commit a crime based on a utility calculation, influenced by a judge who sets a society-wide threshold corresponding to the likelihood of an individual being found guilty and a legislator who sets a society-wide punishment level. We study how the overall crime rate is influenced by the judge's threshold and the legislator's punishment level, propose an objective function for the judge and legislator to minimise, and study the optimal threshold and punishment levels for this objective. We then consider the case in which the overall society is subdivided into multiple groups with varying characteristics, introducing a constraint on fairness in treatment between the groups. We study how an optimal threshold and punishment level might be chosen under this fairness constraint, what ramifications the constraints have on outcomes for individuals, and under what circumstances the constrained optimum agrees with the unconstrained optimum.

1. Introduction

Since at least the work of Becker, the study of crime and punishment as a problem of economics and choice has received significant attention [3–20]. Framed simply, one can conceptualise individuals as rational agents choosing whether or not to commit crimes based on the estimated utility of the two competing choices, while the justice system, by setting standards of proof for conviction and the severity of punishments imposed, attempts to influence this choice in some desired way. Questions then arise as to what the thresholds of evidence or levels of punishments should be that will best optimise some stated objective function; for example, to minimise expected crime rates.

Becker asked how many resources and how much punishment should be used to enforce different kinds of legislation. There, the optimal decisions minimise social loss in income from offences – the sum of damages, costs of apprehension, costs of conviction, and costs of carrying out punishments. This choice of optimal criteria and how it is calculated has led to more discussions. For example, in Becker, the damage to society is the cost of the offence minus the gain to the offender. In other words, the gain for the offender is counted as a gain for society – leading to the conclusion that society should allow efficient crimes. In response, Stigler posits that an offender's illicit gains should not be counted as society's gains, suggesting a change to what the legislators should optimise. This view is not without its problems, since society's decision on what counts as illicit gains changes throughout time [22]. Others have circumvented the issue by focusing on other optimal conditions or narrowing down the details of the illicit activity being considered. For example, Raskolnikov focused on crime where the illicit gains always equal the harm done. Curry and Doyle introduced a voluntary market option for individuals to achieve the same objective as they would with criminal behaviour. With that, Curry and Doyle showed that minimising the cost of crime corresponds with maximising social welfare. The possibilities to be

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

considered are numerous; Polinsky and Shavell give an excellent overview and discussion on deterrence modelling considerations.

At the same time, a large body of work shows that the justice system often has differential effects on various subgroups of the overall population [13–17]. Motivated by this, some recent works on the economics of crime have begun to build various metrics of 'fairness' into the objective function. A fundamental difficulty with this task is that consensus on a formal definition of fairness is lacking. Further, it is known [4, 11] that some fairness metrics are inherently incompatible with others, such that both cannot possibly be achieved simultaneously. And, since these metrics often test the 'outcome' of an algorithm, the issue of fairness is further complicated by the *infra-marginality* problem, where an outcome test for two groups with different risk distributions can suggest bias in favour of one group even if none exists, or sometimes even if the true bias is in favour of the other group [25]. Which definition of fairness is most desirable transcends mathematics and requires moral arguments and philosophical discussions [14, 7].

Despite these caveats about fairness, generally, researchers on the economics of crime simply choose a plausible notion of fairness and proceed from there, cognisant that their choice may not be shared by all. For example, Persico proposed a model that imposes a fairness restriction for the police such that police behaviour is defined as fair when they police two subgroups with the same intensity; this is coupled with the goal of maximising the number of successful inspections. Persico concludes that, under certain conditions, forcing the police to behave more fairly reduces the overall crime rate. As an alternative notion, Jung et al. consider fairness as an equality in conditional false positive rates between groups, and also show that the crime rate is minimised when this constraint is upheld. Our work adopts this same notion of fairness and shares many other traits with Jung et al. However, due to some key differences between our model and that of Jung et al., we find that equalising false positive rates between groups generally does not minimise crime, nor optimise a general objective function, if it is indeed even possible in the first place. However, we show that under certain circumstances and objective function choices, the fair scenario is in fact the global optimiser.

This work aims to accomplish the following. First, to introduce a model of crime and deterrence that, while certainly only a vastly simplified version of reality, includes certain notions not otherwise included in prior models of this type. This is done in Section 2, in which we will discuss Jung et al.'s baseline model and introduce our own, highlighting the key differences. Second, based on our model, we aim to explore how the crime rate of society depends on the various choices of the justice system and to determine how the justice system might 'optimally' achieve a goal of lowering crime while still keeping in mind the negative impact that punishment can have on innocent individuals. This is done in Section 3, where we will explore how the crime rate of a single group reacts with respect to the threshold τ set by the judge and the punishment level κ set by the legislator, and in which we introduce a single group objective function and find optimal values of τ and κ under a variety of circumstances. Finally, to determine based on the model whether or not a specific notion of fairness between groups can be accomplished, and what optimality for the justice system might look like when this fairness consideration for two groups is included in the objective function. This is studied in Section 4, where we extend our analysis to the case of two groups and explore how a notion of fairness impacts the objective function previously introduced in Section 3.

2. Setup and baseline model overview

[1]We start with a version of Jung et al.'s model. Individuals in a societal group k make a binary decision to commit a crime (c) or to remain innocent (i). An individual who chooses to commit a crime receives reward ρ, while an individual who chooses not to commit a crime receives reward ν. The difference between these two quantities is $\gamma \equiv \rho - \nu$, and varies from person to person within the group since

[1] Due to the large number of model parameters, a reference table (Table 1) containing a list of select parameters and their definitions ordered by their appearance in the text has been provided.

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

Table 1. *List of important model parameters and notations with their definition and/or interpretation in plain language ordered by appearance*

Parameter	Definition/interpretation
ρ	P. 3, reward for an individual for committing a crime.
ν	P. 3, reward for an individual for not committing a crime.
γ	P. 3, equals to $\rho - \nu$. Described by probability density Γ_k.
Γ_k	P. 3, probability density of γ for group k.
α_k	P. 3, policing rate for criminals in group k.
β_k	P. 3, policing rate for noncriminals in group k.
s	P. 3, individual's evidence "signal" drawn from distribution θ_k^c or θ_k^i.
θ_k^i	Equation (2.1), distribution of signal for noncriminals in group k. Assumed to be exponential with rate parameter λ_k.
θ_k^c	Equation (2.2), distribution of signal for criminals in group k. Assumed to be exponential with rate parameter ω_k.
λ_k	P. 4, rate parameter for θ_k^i.
ω_k	P. 4, rate parameter for θ_k^c.
p_k	P. 4, equals to $\frac{\omega_k}{\lambda_k} < 1$, distinction between signals of criminals and noncriminals.
τ	P. 4, posterior guilt probability threshold for conviction set by the judge
κ	P. 4, punishment for convicted individuals set by the legislator
FPR_k	P. 5, false positive rate for group k
TPR_k	P. 5, true positive rate for group k
Δ_k	Equation (2.5), also equivalent to $TPR_k - FPR_k$.
C_k	Equation (2.6), crime rate for group k.
N_k	P. 5, population size of group k
χ	P. 6, equal to $\frac{C}{1-C}\frac{1-\tau}{\tau}$, defined for convenience and readability.
f_k	Equation (2.10), defined for convenience and readability.
M	Equation (3.16), objective function for the justice system in the single group case
M_B	Equation (4.13), objective function for the justice system in the two groups case
M_F	Equation (4.15), objective function for the justice system in the two groups case with a fairness notion implemented

opportunities both within and outside of crime might naturally vary; let the density $\Gamma_k(\gamma)$ represent the distribution of this quantity within the group. Clearly, those individuals with $\gamma \le 0$ have no quantitative incentive to commit crime, while those with $\gamma > 0$ do. We will assume throughout that the density $\Gamma_k \ge 0$ is strictly positive at all positive γ values, indicating that it contains some mass in the positive γ region so that there are at least some individuals who are motivated to commit crime.[2]

Each individual within the group may come under suspicion or scrutiny as a possible criminal and be 'investigated' or 'policed'. Let the group-dependent policing rate for those who choose to commit a crime and those who do not be denoted as α_k and β_k, respectively. We will generally assume that $\alpha_k \ge \beta_k$, throughout, in the hope that criminals are at least as likely to be investigated as those who are innocent. Each individual who comes under such scrutiny will produce a random 'signal' $s \ge 0$ that represents effectively the amount of evidence that appears to indicate guilt for that individual. The distribution of signals for criminals and innocents within group k is denoted as $\theta_k^c(s)$ and $\theta_k^i(s)$, respectively. We assume,

[2] This assumption is not key to any of our results, and is certainly not necessary in order for there to be mass at positive γ values. Rather, it is made to slightly simplify some of the analysis in the remainder of the paper, allowing us to ignore certain edge cases that might arise otherwise. Even though this assumption may seem a bit unrealistic, note that even if the true density were identically zero on some interval of γ values, this assumption only requires Γ to be some positive value on this interval, but this value can be chosen as small as one likes, specifically small enough such that the total mass in that region is as close to zero (the true value) as desired.

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

as in Jung et al., that these signals exhibit a Monotone Likelihood Ratio Property (MLRP), meaning that $\frac{\theta_k^c(s)}{\theta_k^i(s)}$ is nondecreasing in s. That is, a higher signal s never denotes a lower likelihood of being guilty vs innocent. For concreteness, throughout the rest of this work, we will assume that the signals s are drawn from exponential distributions

$$\theta_k^i(s) = \lambda_k e^{-\lambda_k s},\ \lambda_k > 0 \tag{2.1}$$

$$\theta_k^c(s) = \omega_k e^{-\omega_k s},\ \omega_k > 0 \tag{2.2}$$

and we define

$$p_k \equiv \frac{\omega_k}{\lambda_k} < 1$$

to guarantee the MLRP. One can interpret the parameter p_k as essentially indicating how 'easy' it is to determine those who are guilty vs innocent within the group k, with p_k close to zero indicating that this determination is relatively easy, and p_k near one indicating that the determination is relatively hard.

Finally, a judge determines whether the evidence indicates guilt, denoted by $z = 1$, or innocence, denoted by $z = 0$, for each individual under scrutiny. In contrast to Jung et al., we focus entirely on the case in which the judge determines guilt versus innocence based on the posterior probability of the individual being a criminal, rather than making the decision based directly on the signal. That is, an individual is classified as guilty if the judge determines that their posterior probability of being a criminal given their signal, their group, and a prior belief on guilt is larger than some threshold $0 < \tau < 1$, i.e. $P(c|s, k) > \tau$; otherwise they are found innocent. The details of this assessment will be provided later. Jung et al. instead focus on policies that are based directly on the signal itself, and which are also based on simple thresholding. There is a discussion within Jung et al. of the possibility of posterior thresholding, in which the authors point out that there is rough equivalency between these two methods – a threshold on signal can be translated to a threshold on posterior and vice versa – as we will elaborate on below. However, for the policy problem that is presented in Jung et al., minimising total crime across groups, it is generally true that the optimal policy will correspond to posterior thresholds that are different for different groups. Here is where our approach fundamentally differs: we will insist in this work that only one posterior threshold exists, τ, and then frame our policy problem under this constraint. We make this choice because one focus of this work is on exploring fairness within the context of this problem, and having separate posterior thresholds for each societal group may violate general ethical and/or legal standards held in many societies; this, despite evidence that in reality these standards are sometimes violated [13–6].

As a final aspect of the model, anyone found guilty receives punishment κ, set by the legislator, regardless of group. As with the common posterior threshold mentioned above, this assumption comports with general ethical and/or legal standards held in many societies, despite evidence that in reality these standards are sometimes violated [13–6]. In practice, legislators often set a range of punishment levels that judges can choose from to assign to individuals found guilty. But, we will later show that in our framework, the legislator is able to find a single optimal punishment level according to a specific objective function. So, throughout this work, we will assume that the legislator only picks one single punishment level κ for the entire society, which could be the optimal one. This assumption could be relaxed to form a model closer to reality, but with more complications, including a more complex calculation on the part of the society members that would include a probability distribution of κ values that they might receive; this may be considered in future work.

Given the above, an individual then chooses to commit a crime if their expected net utility from committing a crime is higher than not committing crime. The inequality governing an individual's decision to commit a crime then becomes

$$\rho - \alpha_k \kappa P(z = 1|c, k) > \nu - \beta_k \kappa P(z = 1|i, k). \tag{2.3}$$

Rearranged, we have

$$\kappa \Delta_k < \rho - \nu = \gamma, \tag{2.4}$$

where

$$\Delta_k \equiv \alpha_k P(z = 1|c, k) - \beta_k P(z = 1|i, k) = TPR_k - FPR_k \tag{2.5}$$

is a measure of the difference in probability of being found guilty for criminals and innocents within group k, and is therefore the difference between the true positive rate TPR_k and false positive rate FPR_k for group k. Then the overall crime rate, measured as the fraction of people choosing to commit crime for group k satisfies

$$C_k = \int_{\kappa \Delta_k}^{\infty} \Gamma_k(\gamma)\, d\gamma. \tag{2.6}$$

We assume that the judge has knowledge of the relevant parameters and distributions for all groups, and can then use this knowledge to aid in determining guilt versus innocence for an individual producing a given signal s. Specifically, the judge makes a Bayesian posterior calculation on the probability of an individual being a criminal based on their signal, group membership, and the known society-wide crime rate C, then decides guilt vs innocence based on the threshold value τ on this posterior probability. Let the society-wide crime rate be computed as

$$C = \sum_{k=1}^{\mathcal{G}} N_k C_k,$$

where $\mathcal{G}$ is the number of distinct societal groups (however these might be defined) and N_k and C_k are the fraction of the total society that belongs to group k and the crime rate of group k, respectively. The posterior probability of an individual being a criminal after observing some signal s given a group k is

$$P(c|s, k) = \frac{\alpha_k \theta_k^c(s) C}{\alpha_k \theta_k^c(s) C + \beta_k \theta_k^i(s)(1 - C)}. \tag{2.7}$$

We note that, from the point of view of an accurate posterior probability, equation (2.7) should use C_k rather than C. However, doing so would essentially represent a prejudice on the part of the judge toward convicting certain groups more readily than others. Since one focus of our work is enforcing fairness within the model, and such a prejudice could readily be construed as unfair, we opt to consider the less biased calculation that uses C rather than C_k.

Recall that an individual is classified as guilty ($z = 1$) if their posterior probability of being a criminal is greater than the threshold set by the judge, $P(c|s, k) > \tau$, with $0 < \tau < 1$. Then, a crime rate of zero would lead to no one ever being found guilty via (2.7). However, in this circumstance, those individuals with $\gamma > 0$ will certainly commit crimes, since they will be guaranteed not to be punished, leading to a contradiction. Hence, it must be the case that $C_k > 0$, i.e., there will always be some non-zero level of crime if individuals with $\gamma > 0$ exist. Because of this and our assumption about the distributions θ, without loss of generality, we can rewrite (2.7) as

$$P(c|s, k) = \left[1 + \frac{\beta_k \theta_k^i(s)}{\alpha_k \theta_k^c(s)} \frac{(1 - C)}{C}\right]^{-1}. \tag{2.8}$$

Due again to the MLRP for the distributions θ_k^i and θ_k^c, for any given crime rate $C < 1$ the posterior probability of guilt increases with increasing signal, approaching unity as $s \to \infty$ and having a minimum value at $s = 0$; for $C = 1$ the probability is always (correctly) unity for any signal. Then for any $C < 1$ and decision threshold τ, there will always exist a threshold signal value s_τ such that only those individuals

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

with $s \geq s_\tau$ are found guilty. If $\tau \leq P(c|0,k) \equiv \tau_0$ this threshold is just $s_\tau = 0$: everyone is always found guilty in this case. Otherwise, the signal threshold is given by

$$s_\tau = -\frac{1}{\lambda_k - \omega_k} \ln\left(\frac{p_k \alpha_k}{\beta_k}\chi\right), \tag{2.9}$$

where

$$\chi \equiv \frac{C}{1-C}\frac{1-\tau}{\tau}.$$

To reiterate, whenever

$$f_k \equiv \frac{p_k \alpha_k}{\beta_k}\chi = \frac{p_k \alpha_k}{\beta_k}\frac{C}{1-C}\frac{1-\tau}{\tau} \leq 1 \tag{2.10}$$

then the signal threshold is given by (2.9); otherwise we are in the regime where $\tau \leq \tau_0$ and everyone is found guilty for all signals, so that $s_\tau = 0$ and we can effectively set $f_k = 1$. The existence of a unique s_τ for each τ highlights the equivalency between posterior and signal thresholding mentioned above.

We can now compute the total probability of being found guilty conditional on investigation for both criminals and innocents of group k:

$$P(z=1|i,k) = P(s \geq s_\tau|i,k) = \int_{s_\tau}^{\infty} \theta_k^i(s)\, ds = e^{-\lambda_k s_\tau} = [\min(f_k, 1)]^{\frac{1}{1-p_k}} \tag{2.11}$$

$$P(z=1|c,k) = P(s \geq s_\tau|c,k) = \int_{s_\tau}^{\infty} \theta_k^c(s)\, ds = e^{-\omega_k s_\tau} = [\min(f_k, 1)]^{\frac{p_k}{1-p_k}}. \tag{2.12}$$

Multiplying the quantities in (2.11) and (2.12) by β and α, respectively, gives the false positive rate FRP_k and true positive rate TPR_k for group k. Plugging equations (2.11) and (2.12) into equation (2.5), and being careful of inequality (2.10) we have

$$\Delta_k = \begin{cases} \alpha_k f_k^{\frac{p_k}{1-p_k}} - \beta_k f_k^{\frac{1}{1-p_k}}, & f_k \leq 1 \\ \alpha_k - \beta_k, & f_k > 1. \end{cases} \tag{2.13}$$

We observe that since $0 < p_k < 1$, $\frac{p_k}{1-p_k} < \frac{1}{1-p_k}$. So, $f_k^{\frac{p_k}{1-p_k}} \geq f_k^{\frac{1}{1-p_k}}$ when $f_k \leq 1$. Combined with our assumption that $\alpha_k \geq \beta_k$ we have $TPR_k \geq FPR_k$ and $\Delta_k \geq 0$.

Given all of the group-dependent parameters, as well as τ and κ, (2.6) is an implicit equation for the crime rate C_k. In the remainder of this work, we study solutions to this equation in both the single group and two-group cases, and use the crime rates obtained to solve optimisation problems for τ and κ.

3. One group

In this section, we will focus on the properties of solutions to equation (2.6) for one single group. For ease of reading, we drop the subscript k since in this case all parameters and variables belong to a singular group, and in this case, C and C_k are identical. We define two functions based on the LHS and RHS of equation (2.6):

$$g(C) = C \tag{3.1}$$

$$h(C) = \int_{\kappa\Delta(C)}^{\infty} \Gamma(\gamma)\, d\gamma. \tag{3.2}$$

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

In (3.2) we have emphasised that Δ is a function of C. The crime rate solves $g = h$. The function g is straightforward, but we will need to explore the function h further. First, we will show that for some parameter region, h has a minimum in the interior of its domain as described in the following lemma.

Lemma 3.1. *$h(C)$ has a minimum at $C_* = \tau$ when $\frac{p\alpha}{\beta} < 1$. The minimum is*

$$h(C_*) = \int_{\kappa\Delta_*}^{\infty} \Gamma(\gamma)\, d\gamma$$

where

$$\Delta_* = \left(\alpha \left(\frac{p\alpha}{\beta}\right)^{\frac{p}{1-p}} - \beta \left(\frac{p\alpha}{\beta}\right)^{\frac{1}{1-p}} \right)$$

is a constant.

Proof. Recall that f is a function of C given in (2.10). Then equation (2.13) is equivalent to

$$\Delta = \begin{cases} \alpha f^{\frac{p}{1-p}} - \beta f^{\frac{1}{1-p}}, & C < \dfrac{1}{1 + p\frac{\alpha}{\beta}\frac{1-\tau}{\tau}} \\[2ex] \alpha - \beta, & C \geq \dfrac{1}{1 + p\frac{\alpha}{\beta}\frac{1-\tau}{\tau}}, \end{cases} \tag{3.3}$$

where f is still a function of C. And so, when $C > \frac{1}{1+p\frac{\alpha}{\beta}\frac{1-\tau}{\tau}} \equiv C_0$,

$$h(C) = h(C_0) = \int_{\kappa(\alpha-\beta)}^{\infty} \Gamma(\gamma)\, \mathrm{d}\gamma, \tag{3.4}$$

a constant with respect to C. Note that

$$h(0) = \int_{0}^{\infty} \Gamma(\gamma)\, \mathrm{d}\gamma \tag{3.5}$$

is also a positive constant. The derivative of h is computed to be

$$\frac{\mathrm{d}h}{\mathrm{d}C} = \begin{cases} -\Gamma(\kappa\Delta)\kappa \dfrac{\mathrm{d}\Delta}{\mathrm{d}C} = -q(\alpha f^{-1} p - \beta), & C < C_0 \\[2ex] 0, & C > C_0. \end{cases} \tag{3.6}$$

where

$$q \equiv \Gamma(\kappa\Delta)\kappa \frac{1}{1-p} \frac{1}{C(1-C)} f^{\frac{1}{1-p}} > 0. \tag{3.7}$$

Then, any critical point C_* of $h(C)$ in $(0, C_0)$ can only occur when $\alpha f^{-1} p - \beta = 0$. Substituting f from (2.10) and solving, we find the critical point $C_* = \tau$, which is also equivalent to $\chi = 1$. However, this critical point will exist iff $C_* < C_0$; after some algebra this is equivalent to

$$\frac{p\alpha}{\beta} < 1. \tag{3.8}$$

Under this condition, we have that for $C < C_*$, $\frac{dh}{dC} < 0$, while for $C_* < C < C_0$ we have $\frac{\mathrm{d}h}{\mathrm{d}C} > 0$, indicating a minimum at $C = C_*$. Further,

$$h(C_*) = \int_{\kappa\Delta_*}^{\infty} \Gamma(\gamma)\, \mathrm{d}\gamma \tag{3.9}$$

where

$$\Delta_* = \left(\alpha \left(\frac{p\alpha}{\beta} \right)^{\frac{p}{1-p}} - \beta \left(\frac{p\alpha}{\beta} \right)^{\frac{1}{1-p}} \right) \tag{3.10}$$

is a constant. □

The proof above also leads to the following lemma:

Lemma 3.2. *$h(C)$ is monotonically decreasing on $(0, C_0)$ when $\frac{p\alpha}{\beta} \geq 1$.*

Next, we will prove a lemma that will help us understand how many times g and h intersect which corresponds to the number of solutions to the crime rate equation.

Lemma 3.3. *$\frac{dh}{dC} < 1$ when $1 - \frac{1}{q\beta} \leq \frac{p\alpha}{\beta}$ for $C \in (0, C_0)$.*

Proof. For ease of reading we define $\phi(C) = \alpha f^{-1} p - \beta = \beta \frac{1-C}{C} \frac{\tau}{1-\tau} - \beta$. We first compute

$$\frac{\mathrm{d}\phi}{\mathrm{d}C} = \frac{\mathrm{d}}{\mathrm{d}C} \left(\beta \frac{1-C}{C} \frac{\tau}{1-\tau} \right) \tag{3.11}$$

$$= -\beta \frac{1}{C^2} \frac{\tau}{1-\tau} < 0. \tag{3.12}$$

So, in the domain $C \in (0, C_0)$,

$$\phi(C_0) < \phi(C) \tag{3.13}$$

Here, $\phi(C_0) = p\alpha - \beta$. So, by rearranging our initial assumption and since $f^{-1} > 1$,

$$-\frac{1}{q} \leq p\alpha - \beta < \alpha f^{-1} p - \beta. \tag{3.14}$$

Rearranging, we have

$$\frac{\mathrm{d}h}{\mathrm{d}C} = -q(\alpha f^{-1} p - \beta) < 1. \tag{3.15}$$

□

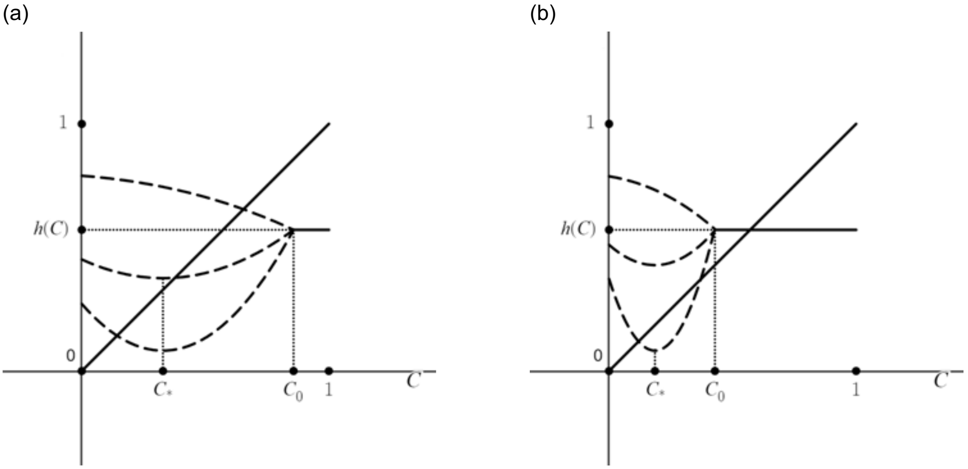
The lemmas above lead to the following uniqueness theorem.

Theorem 3.4. *If $1 - \frac{1}{q\beta} \leq \frac{p\alpha}{\beta}$ for $C \in (0, C_0)$ then there is a unique solution to equation (2.6).*

Proof. Recall from equation (3.5) that $0 < h(0) \leq 1$. Also, $h(1) < h(0) \leq 1$ from properties of cumulative distribution functions and our assumption that $\Gamma > 0$. We have $g(0) = 0$ and $g(1) = 1$. If $\frac{p\alpha}{\beta} \geq 1$, then $h(c)$ is monotonic by Lemma 3.2. And so there must only be one intersection between $g(C)$ and $h(C)$. If $1 - \frac{1}{q\beta} \leq \frac{p\alpha}{\beta} < 1$ for $C \in (0, C_0)$, then by Lemma 3.3, $\frac{dh}{dC} < 1 = \frac{dg}{dC}$. Similarly, there must be only one intersection between $g(C)$ and $h(C)$. □

However, if it is the case that $\frac{p\alpha}{\beta} < 1 - \frac{1}{q\beta}$ for some $C \in (0, C_0)$, there could be multiple solutions, as illustrated in Figure 1. To summarise the findings of the theorems above and as illustrated in Figure 1, when $\frac{p\alpha}{\beta} \geq 1$ there is a unique crime rate for any given τ and κ, and the crime rate is increasing with τ for any $\tau > \tau_0$; that is, crime is minimised when everyone investigated is found guilty. On the other hand, when $\frac{p\alpha}{\beta} < 1$, there are potentially multiple crime rates consistent with (2.6) for certain τ and κ. However, there is a threshold $\tau_* = C_* = h(C_*)$ that yields the smallest crime rate possible, in which not everyone investigated is found guilty.

Recall that $p = \frac{\omega}{\lambda} < 1$ is the ratio of the decay rate of the signal for innocents and the decay rate of the signal for criminals, and that p closer to 1 indicates that the signal of the criminals and innocents are less distinguishable while lower p indicates that the signal of criminals and innocents are more distinguishable. Similarly, the ratio $\frac{\beta}{\alpha}$ denotes how different the policing rate is for innocents vs. criminals, with a



One intersection

Multiple intersections

Figure 1. *Intersection(s) of $g(C)$ and $h(C)$. The dashed lines from top to bottom correspond to the regions $\frac{p\alpha}{\beta} \geq 1$, $1 - \frac{1}{qb} < \frac{p\alpha}{\beta} < 1$, and $\frac{p\alpha}{\beta} < 1 - \frac{1}{qb}$, respectively. The solid horizontal line indicates the region in which $f = 1$. Panel (a) is constructed at a lower threshold τ than that used in (b).*

low ratio meaning that innocents are policed notably less than criminals and a ratio approaching 1 indicating that they are policed at essentially the same rate. Let us rewrite the region $\frac{p\alpha}{\beta} < 1$ instead as $p < \frac{\beta}{\alpha}$. Then this region corresponds to one in which the signals between the innocents and the criminals are more distinguishable than their policing rate ratio would suggest. Intuitively, this means that the judge is able to add something valuable to the guilt determination process, and further narrow down who is guilty vs innocent beyond the relatively crude distinction made by the police. Because of this, the judge can choose a threshold τ_* that minimises crime rate without finding everyone guilty, as shown above. Conversely, the region $\frac{p\alpha}{\beta} \geq 1$, equivalently $p \geq \frac{\beta}{\alpha}$, corresponds to the case where the signals between the innocents and criminals are at most as distinguishable as their policing rate ratio would suggest. In other words, the judge is not as efficient as the police in determining guilt vs innocence, and does not add much of value to the process. Here, the judge minimises crime by just finding everyone guilty, as in this case, the police will have already largely been able to determine guilt before any trial occurs. Generally, it seems plausible that the former case, where the judge is better able to determine guilt vs innocence than the police, is the more realistic of the two.

In the next two subsections, we will discuss the case where there could be multiple solutions. Then, we will explore how the crime rate changes in relation to the threshold τ and the punishment level κ and consider a reasonable objective function for the judge and the legislator to minimise.

3.1. Remark on multiple crime rates

As described above, equation (2.6) may be consistent for several different crime rates C, that is, the solution might not be unique in some parameter regions. In that case, it is not immediately clear which of these consistent crime rates would manifest, as the overall model assumes that criminals can determine their expected utilities perfectly, and those depend on the crime rate observed. However, we will show that, by our assumption that individuals want to maximise their utility, the smallest of these consistent crime rates will be the one to occur.

Suppose there are multiple solutions to (2.6), labelled $C^{(1)} < C^{(2)} < \cdots < C^{(n)}$. Each such crime rate can be thought of as representing a Nash Equilibrium of the system. That is, any solution $C^{(i)}$ corresponds to a value $\gamma^{(i)}$ that has two properties: $C^{(i)} = \int_{\gamma^{(i)}}^{\infty} \Gamma(\gamma) \mathrm{d}\gamma$ (the solution is consistent); and if individuals with $\gamma > \gamma^{(i)}$ commit crimes and those with $\gamma < \gamma^{(i)}$ do not, no individual is tempted to deviate from

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

this unilaterally (a Nash Equilibrium). Note that these threshold γ values lie in the order $\gamma^{(n)} < \gamma^{(n-1)} < \cdots < \gamma^{(1)}$.

However, the expected utilities for individuals among these Nash Equilibria are not equal. Note that those with $\gamma > \gamma^{(1)}$ will commit crimes no matter which equilibrium is selected, and those with $\gamma < \gamma^{(n)}$ will not commit crimes no matter which is selected, so that we need only consider those individuals with $\gamma^{(n)} < \gamma < \gamma^{(1)}$ and determine which equilibrium they might prefer. For a fixed behaviour – commit crimes vs not – utility is decreasing with increasing probability of punishment, which itself increases with crime rate. Hence, for any given individual under consideration we need only contrast two of the equilibria: $C^{(1)}$, in which case they do not commit crime and crime is as low as possible; and $C^{(j)}$, which is the lowest crime rate for which the corresponding $\gamma^{(j)}$ is less than the γ value of the individual in question, which is the lowest crime state in which that person commits crime. Then this individual will prefer the equilibrium at $C^{(1)}$ so long as

$$\nu - \beta\kappa P\left(z = 1|i; C^{(i)}\right) > \rho - \alpha\kappa P\left(z = 1|c; C^{(j)}\right).$$

Upon rearranging terms and writing existing quantities in terms of γ_1, the above inequality is equivalent to

$$\gamma < \gamma^{(1)} + \alpha\kappa\left[P\left(z = 1|c; C^{(j)}\right) - P\left(z = 1|c; C^{(1)}\right)\right].$$

Noting that the term in brackets is positive, this inequality holds for all individuals in question, meaning that they all prefer the lowest crime equilibrium above all others. Therefore, even in cases where there are multiple solutions to (2.6), one would expect that the lowest crime solution should be the one obtained.

3.2. *Judge's and legislator's objective*

We now model the choice of τ and κ as optimisation problems for the judge and legislator. As a very simple first possibility, perhaps the judge and legislator are working in unison to simply minimise the crime rate C. Given the results above, for a fixed punishment level κ, crime is minimised in one of two ways. First, if $\frac{\rho\alpha}{\beta} \geq 1$, then crime is minimised at rate $C_0 = h(C_0)$ from (3.4), since $h(C)$ is monotonically decreasing on $(0, C_0)$ in this case. This corresponds to selecting any threshold $\tau \in (0, \tau_0]$, in which case all individuals are found guilty, and $\tau_0 = 1/\left[1 + \frac{\beta}{\rho\alpha}\frac{1-C_0}{C_0}\right]$. If $\frac{\rho\alpha}{\beta} < 1$, then crime is minimised at rate $C_* = h(C_*)$ from (3.9), since that is where $h(C)$ is minimised in this case (and cases of multiple solutions here will still exhibit the smallest crime rate possible). This corresponds to selecting threshold $\tau = \tau_* = C_*$; in this case, not all individuals are found guilty. In either of these cases, the crime rate simply decreases with κ, indicating that arbitrarily large punishment should be sought, and no global minimum of crime truly exists.

Of course, it is not clear that in reality severe punishment (for a given threshold) always leads to lower crime. This model makes many simplifying assumptions that are perhaps only approximately true, not least of which the idea that the choice to commit crimes is always based on a logical cost–benefit analysis; crimes of passion do exist, after all. This assumption could be relaxed to allow for individuals to commit crimes even if it is not economically beneficial, but we leave that for future work. But beyond this, even if the model were a perfect simulacrum of the criminal justice system, the objective of simply minimising crime does not appear very satisfying. It suggests draconian punishment, ignoring the fact that these punishments are sometimes, unfortunately, meted out to innocent people. Further, it indicates that the posterior threshold for punishment should be very small – either small enough so that all are guilty, or set to match C_*, which is being made as small as possible – which generally conflicts with Western ideals that posteriors at least be "the preponderance of evidence," if not higher.

We therefore propose an alternative objective function. Certainly, low crime is still desired, as that minimises the impact of crime on victims, among many other things. However, this desire should be balanced against harm that could also be done to innocent individuals through erroneous punishment. The portion of the objective indicating harm could take many forms and include many items, such as

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

opportunity costs incurred by punishing innocent individuals, or fear and distrust that very severe punishments might instil in citizens. However, to avoid introducing too many other considerations to the model, we stick with a relatively basic term that can be included without much difficulty, and which essentially directly measures the amount of punishment applied to innocent people. We therefore propose that the judge and legislator might consider the objective function

$$M = C + \lambda\kappa^n(1 - C)FPR. \tag{3.16}$$

Here, $\lambda > 0$ and $n > 0$ are both parameters that change the balance between desiring low crime vs low punishment for innocent individuals who are found guilty. We note that for $n = 1$, a seemingly natural choice, the second term is directly proportional to the total amount of punishment meted out to innocents, with a constant of proportionality λ. However, as we will show below, choosing $n = 1$ leads to an unsatisfying solution to the optimisation problem.

While the natural choice of parameters over which to optimise M are κ and τ, it is easier to analyse the system by choosing to parameterise with κ and χ. This parametrisation is equivalent, so long as we note that some κ, χ combinations may not be feasible. Specifically, any κ, χ combination has a well-defined crime rate, and that crime rate, when combined with χ gives a well-defined τ. However, for a given κ, there could exist two (or more) values $\chi_1 < \chi_2$, with corresponding $C_1 < C_2$ that both give the same τ. As noted above, in such cases when a single τ gives rise to multiple crime rates, only the lowest rate is realisable, hence the combination κ, χ_2 is not feasible. While this could potentially be a problem moving forward, we note that this issue will not arise when $\frac{p\alpha}{\beta} \geq 1$, and even when $\frac{p\alpha}{\beta} < 1$, we can be sure that any κ with $\chi \leq 1$ is feasible. This is because, when a single τ could yield multiple crime rates via (2.6), either one of those crime rates has $\chi \leq 1$ and the others have $\chi > 1$, and the $\chi \leq 1$ rate is the feasible one, or all of the crime rates have $\chi > 1$. We will revisit this point later.

Seeking critical points of M gives the following equations

$$\frac{\partial M}{\partial \kappa} = -\Gamma(\kappa\Delta)\Delta(1 - \lambda\kappa^n FPR) + \lambda n\kappa^{n-1}(1 - C)FPR = 0 \tag{3.17}$$

$$\frac{\partial M}{\partial \chi} = \left[-\Gamma(\kappa\Delta)\kappa\frac{\partial\Delta}{\partial FPR}(1 - \lambda\kappa^n FPR) + \lambda\kappa^n(1 - C)\right]\frac{\partial FPR}{\partial\chi} = 0. \tag{3.18}$$

After some algebra, one can show that in order for there to exist a simultaneous solution to the equations above, it must be the case that

$$\chi = \frac{n - \frac{1}{p}}{n - 1} \equiv \chi_o. \tag{3.19}$$

Note that since $\chi > 0$, equation (3.19) requires either that $n < 1$ or that $n > \frac{1}{p} > 1$ in order for such a critical point to exist. The $n < 1$ case always leads to $f > 1$ and therefore should not be considered further. Assuming $n > \frac{1}{p}$, then $\chi_o < 1$, which means this is certainly a feasible point. However, we must still determine if $f = \frac{p\alpha}{\beta}\chi_o \leq 1$. This is automatically the case when $\frac{p\alpha}{\beta} < 1$; that is, when crime has a minimum at C_*. However, if $\frac{p\alpha}{\beta} \geq 1$, then this requirement places an upper bound on n for the existence of the critical point, such that $n < \frac{\alpha-\beta}{\alpha p-\beta}$.

Assume for now that all requirements are met for a physically relevant χ_o. Let FRP_o and TPR_o be the false positive and true positive rates obtained when $\chi = \chi_o$, and let $\Delta_o = TPR_o - FPR_o$. Then any interior critical points are located at (κ_o, χ_o), where κ_o satisfies

$$-\Gamma(\kappa_o\Delta_o)\Delta_o(1 - \lambda\kappa_o^n FPR_o) + \lambda n\kappa_o^{n-1}[1 - C(\kappa_o, \Delta_o)]FPR_o = 0, \tag{3.20}$$

which in general would have to be solved numerically for κ_o. But, by considering the behaviour of the above expression as $\kappa \to 0$ and $\kappa \to \infty$, we note that the above equation can be made to hold true at any $\kappa_o > 0$ by a careful choice of λ. Further, these λ values become arbitrarily large as $\kappa_o \to 0$ and arbitrarily small (assuming boundedness of Γ at large arguments) as $\kappa_o \to \infty$, so that any choice of λ should yield at least one solution κ_o.

We now switch from χ to FPR, and check the boundaries of the domain, $\kappa \in [0, \infty)$ and $FPR \in [0, \beta]$, to see whether the critical point(s) above are the only possibilities for a global minimum or not. If $n < \frac{1}{p}$, the next few lines will show that $M \to 0$ as $\kappa \to \infty$ in a particular way, leaving no global minimum. To get $M \to 0$, we need both C and $\lambda\kappa^n(1-C)FPR$ tending to zero. As a result, we need $\kappa^n FPR \to 0$. For this to happen as $\kappa \to \infty$, we need $FPR \to 0$. Now note that $TPR = \frac{\alpha}{\beta^p} FPR^p$. Since $FPR \to 0$ and $p < 1$, $FPR^p \gg FPR$. And so, for the requirement that $C \to 0$,

$$C = \int_{\kappa\Delta}^{\infty} \Gamma(\gamma)\,\mathrm{d}\gamma = \int_{\kappa(\frac{\alpha}{\beta^p} FPR^p - FPR)}^{\infty} \Gamma(\gamma)\,\mathrm{d}\gamma \to 0, \tag{3.21}$$

we need $\kappa(\frac{\alpha}{\beta} FPR^p - FPR) \to \infty$. In particular, we need $\kappa FPR^p \to \infty$. Equivalently, we need $(\kappa^{\frac{1}{p}} FPR)^p \to \infty$ which is true iff $\kappa^{\frac{1}{p}} FPR \to \infty$. Rearranged,

$$\kappa^{\frac{1}{p}-n}(\kappa^n FPR) \to \infty. \tag{3.22}$$

So, since $\kappa^n FPR \to 0$, we need $\kappa^{\frac{1}{p}-n} \to \infty$ in a way that satisfies (3.22). This condition can only be satisfied if $n < \frac{1}{p}$, as claimed. As previously discussed, arbitrarily large punishments are not generally feasible, nor desired, so we will focus on the case where $n > \frac{1}{p}$, where $M \to \infty$ as $\kappa \to \infty$ so long as $FPR \neq 0$.

When $FPR = 0$, we have $TPR = 0$. In turn, $\Delta = 0$ and therefore $\kappa\Delta = 0$. So,

$$M = C_{max} \equiv \int_0^{\infty} \Gamma(\gamma)\,\mathrm{d}\gamma. \tag{3.23}$$

Similarly when $\kappa = 0$, $\kappa\Delta = 0$ and so, $C = C_{max}$ and

$$M = C_{max}. \tag{3.24}$$

We note that by (3.17), when $\kappa = 0$ and $FRP > 0$, $\frac{\partial M}{\partial \kappa} < 0$, indicating that $M = C_{max}$ cannot be the global minimum. Similarly, by (3.18), as $FPR \to 0$ and $\kappa > 0$, $\frac{\partial M}{\partial FPR} < 0$.

When $FPR = \beta$ then $TPR = \alpha$ and $\Delta = \alpha - \beta$, and there is at least one point $\kappa = \kappa_\beta$ that could be a local minimum, which satisfies

$$\frac{\partial M}{\partial \kappa} = -\Gamma(\kappa_\beta\Delta)\Delta(1 - \lambda\kappa_\beta^n\beta) + \lambda n\kappa_\beta^{n-1}[1 - C(\kappa_\beta, \Delta)]\beta = 0. \tag{3.25}$$

At the same time, at (κ_β, β),

$$\frac{\partial M}{\partial FPR} = \lambda\kappa_\beta^n[1 - C(\kappa_\beta\Delta)]\left(-\frac{\alpha p - \beta}{\alpha - \beta}n + 1\right). \tag{3.26}$$

Note then that if $\frac{p\alpha}{\beta} < 1$, the case in which crime is minimised at C_*, $\frac{\partial M}{\partial FPR} > 0$ at (κ_β, β), and so this point is not a minimum. Alternatively, if $\frac{p\alpha}{\beta} \geq 1$, then $\frac{\partial M}{\partial FPR} > 0$ and (κ_β, β) is not a minimum when $n < \frac{\alpha-\beta}{\alpha p-\beta}$, which corresponds to the case in which the critical point(s) (κ_o, χ_o) exist. The final possibility is that $\frac{p\alpha}{\beta} \geq 1$ and $n > \frac{\alpha-\beta}{\alpha p-\beta}$, in which case there are no critical point(s) (κ_o, FPR_o) and there is a local minimum for at least one point (κ_β, β).

The above arguments lead to the following theorem:

Theorem 3.5. *When $n > \frac{1}{p}$ and either $\frac{p\alpha}{\beta} < 1$ or $n < \frac{\alpha-\beta}{\alpha p-\beta}$ there is a global minimum of M in the interior of the domain at (κ_o, FPR_o). When $n > \frac{1}{p}$, $\frac{p\alpha}{\beta} \geq 1$, and $n > \frac{\alpha-\beta}{\alpha p-\beta}$ there is a global minimum of M along the boundary at (κ_β, β). When $n < \frac{1}{p}$, the objective function $M \to 0$ as $\kappa \to \infty$ and $FPR \to 0$ in a particular way and M has no strict global minimum.*

The above results indicate that a wide array of "optimal" justice systems occur based on the choice of n and λ in (3.16). For example, a system with low C and high τ is optimal with a very small value of χ_o coupled with a somewhat large κ_o, indicating an n only slightly above $1/p$ and a relatively small λ. Alternatively, as $n \to \infty$, $\chi \to 1$. This χ corresponds to choosing τ_* that solely minimises the crime rate as in Lemma 3.1, in which $C_* = \tau$. In other words, as the scaling of M with κ grows ever larger, the

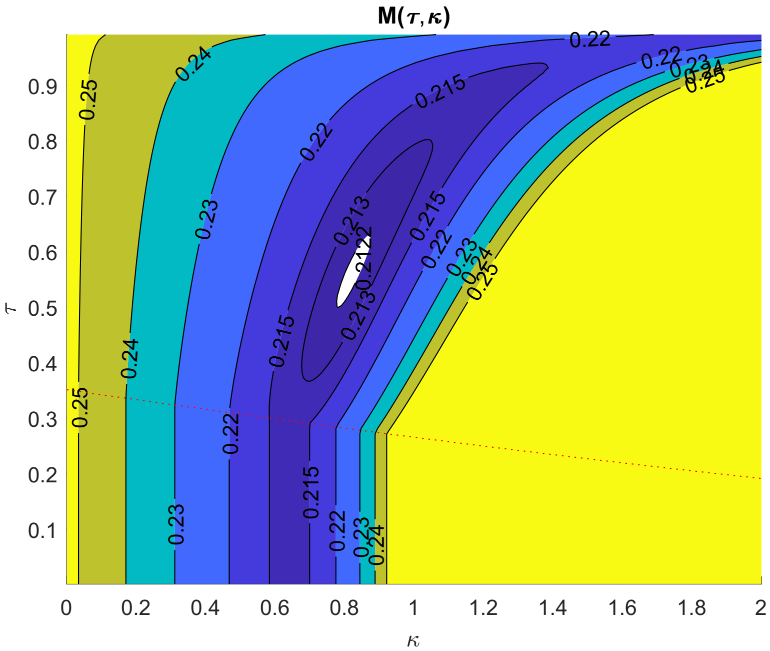


Figure 2. *Contour plot of $M(\kappa, \tau)$ with parameters $\lambda = 1, n = 6, \alpha = 0.8, \beta = 0.2$ and $p = 0.2$. Here $\gamma \sim \mathcal{N}(-2, 3)$. This figure shows a minimum for M at $(\tau, \kappa) = (0.56, 0.82)$. The red dotted line plots τ_0 for each fixed κ.*

judge's best course of action is in solely minimising the crime rate C, and in this case it is not possible to have both a low C and a high τ.

Figure 2 is a contour graph that shows the behaviour of the objective function M in a numerical simulation. To construct this plot, for each threshold τ and punishment level κ indicated, we solve the crime rate equation (equation (2.6)) and then compute the objective function M. In choosing the parameter values for the numerical simulation, we attempt to keep in mind what a reasonable society might have, though admittedly, the parameters chosen are not based on any empirical values. We choose $p = 0.2$ for our simulation, a case where the signals of the criminals and innocents are relatively easy to distinguish. We set the policing rate of criminals at $\alpha = 0.8$. Meanwhile, we set the policing rate of the innocents to be lower, at $\beta = 0.2$ – partially to set $\frac{p\alpha}{\beta} < 1$, where the judge can minimise crime rate without finding all investigated individuals guilty. We choose $\gamma \sim \mathcal{N}(-2, 3)$, giving a population in which the maximum possible crime rate is $\int_0^\infty \mathcal{N}(1, 3) = 0.25$. We choose $n = 6$ to satisfy the condition of a global minimum in the interior as in Theorem 3.5. We choose the parameter λ to be 1 for simplicity. We found that $(\tau, \kappa) = (0.56, 0.82)$ minimises M with value of 0.21; the crime rate is 0.20 and the *FPR* is 0.02. In comparison, with the same parameters, and when $\kappa = 0.82$ is fixed, as in Lemma 3.1, the judge can minimise the crime rate even further to 0.19 with a lower threshold $\tau = 0.19$, but with a higher $FPR = 0.18$.

4. Two groups

Having covered several aspects of the model in the context of a single population, we now turn back to the possibility of multiple groups, specifically with notions of fairness between groups in mind. In the fairness in algorithms literature, one definition of fairness that has been used is based on *disparate impact*: when the outcome of the algorithm disproportionately harms or helps specific social groups vs others. Within the context of criminal justice, one particular notion of fairness that removes a form of disparate impact is a requirement of parity of false positive (or negative, depending on context) rates

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

across groups. That is, if a false positive refers, as it does in this manuscript, to convicting an innocent person, then one might want a fair "algorithm" to make sure that different social groups within the population all suffer this at the same (ideally low) rate.

To include this notion of fairness in our discussion, we explore our model with two groups, denoted simply as groups 1 and 2. For comparison, we will first run the same analysis as in the one-group case without including the notion of fairness. In this case, the total crime rate C that appears in (2.7) is

$$C = N_1C_1 + N_2C_2 = N_1\int_{\kappa\Delta_1}^{\infty}\Gamma_1(\gamma)\,\mathrm{d}\gamma + N_2\int_{\kappa\Delta_2}^{\infty}\Gamma_2(\gamma)\,\mathrm{d}\gamma, \tag{4.1}$$

where N_1 and N_2 are proportions of the total population for groups 1 and 2, respectively. For the two-group model to be meaningfully different from the one-group model, we require that the two groups differ in their values for p and/or α, β. This is due to our restriction above that the crime rate used by the judge to determine guilt is the overall societal crime rate. Hence, if the two groups shared identical values for p, α, and β, then the lower bounds for each of the two integrals in (4.1) would be identical, and the two integrands could be combined into an overall societal distribution of Γ, leaving the one-group problem.

With similar proof ideas as in the previous section, one can show that, assuming a fixed value for the punishment level κ, the lowest possible crime rate occurs at a well-defined value for τ. We now define $h(C)$ to be the RHS of equation (4.1). Without loss of generality, let group 1 be the group with the higher value of $p_k\alpha_k/\beta_k$; for any given τ and C, we then have $f_1 > f_2$. For a fixed $\tau, f_1 = 1$ will be achieved at a lower crime rate than needed for $f_2 = 1$. Then we have the following theorem, which can be generalised to any number of finite groups:

Theorem 4.1. *Assume κ is constant. When $\frac{p_1\alpha_1}{\beta_1} < 1$, C has a minimum of*

$$C_* = h(C_*) = N_1\int_{\kappa\Delta_{1*}}^{\infty}\Gamma_1(\gamma)\,d\gamma + N_2\int_{\kappa\Delta_{2*}}^{\infty}\Gamma_2(\gamma)\,d\gamma$$

where

$$\Delta_{k*} = \left(\alpha_k\left(\frac{p_k\alpha_k}{\beta_k}\right)^{\frac{p_k}{1-p_k}} - \beta_k\left(\frac{p_k\alpha_k}{\beta_k}\right)^{\frac{1}{1-p_k}}\right),$$

at $\tau = \tau_* = C_*$ ($\chi = 1$). When $\frac{p_2\alpha_2}{\beta_2} < 1$ but $\frac{p_1\alpha_1}{\beta_1} > 1$, C has a minimum of

$$C_m = h(C_m) = N_1\int_{\kappa(\alpha_1-\beta_1)}^{\infty}\Gamma_1(\gamma)\,d\gamma + N_2\int_{\kappa\Delta_{2*}}^{\infty}\Gamma_2(\gamma)\,d\gamma,$$

at $\tau = \tau_m = C_m$ ($\chi = 1$). When $\frac{p_2\alpha_2}{\beta_2} > 1$, C has a minimum of

$$C_0 = h(C_0) = N_1\int_{\kappa(\alpha_1-\beta_1)}^{\infty}\Gamma_1(\gamma)\,d\gamma + N_2\int_{\kappa(\alpha_2-\beta_2)}^{\infty}\Gamma_2(\gamma)\,d\gamma,$$

for any $\tau \leq \frac{1}{1+\frac{p_1\alpha_1}{\beta_1}\frac{C_0}{1-C_0}}$.

The theorem above already illustrates the possibility of a large disparity in impact between the two groups if crime is simply minimised. The clearest case is when $\frac{p_2\alpha_2}{\beta_2} < 1$ but $\frac{p_1\alpha_1}{\beta_1} > 1$, in which case minimising crime means finding all investigated members of group 1 guilty, while some investigated members of group 2 are set free.

4.1. Enforcing parity of false positive rates

We now consider what would be required to equalise false positive rates between groups 1 and 2. The false positive rate for group k is

$$FPR_k = \beta_k P(z=1|i) = \begin{cases} \beta_k f_k^{\frac{1}{1-p_k}}, & f_k < 1 \\ \beta_k, & f_k \geq 1. \end{cases} \tag{4.2}$$

In the region where $f_2 \geq 1$, $FPR_1 = \beta_1$ and $FPR_2 = \beta_2$; all investigated persons are found guilty in this regime. In other words, the false positive rates for both groups equalise if and only if their investigation rates for innocents are equal, i.e. $\beta_1 = \beta_2$. When $f_2 < 1$ but $f_1 \geq 1$, the false positive rates for both groups equalise if and only if

$$\beta_1 = \beta_2 f_2^{\frac{1}{1-p_2}}. \tag{4.3}$$

Rearranging,

$$\chi = \frac{\beta_2}{p_2\alpha_2}\left(\frac{\beta_1}{\beta_2}\right)^{1-p_2} \equiv \chi_f. \tag{4.4}$$

However, for (4.4) to be valid, $\chi = \chi_f$ must indeed lead to $f_2 < 1$ and $f_1 \geq 1$. This implies

$$\frac{\beta_1}{\beta_2} < \min\left[1, \left(\frac{p_1\alpha_1}{p_2\alpha_2}\right)^{\frac{1}{p_2}}\right].$$

Note that this requirement is mutually exclusive of the one above.

In the region where $f_1 \leq 1$, the false positive rates for both groups equalise if and only if

$$\beta_1 f_1^{\frac{1}{1-p_1}} = \beta_2 f_2^{\frac{1}{1-p_2}}. \tag{4.5}$$

If $p_1 = p_2 = p$, this will occur for any χ, but only if $\frac{\beta_1}{\beta_2} = \left(\frac{\alpha_1}{\alpha_2}\right)^{1/p}$. Otherwise, we rearrange to find

$$\chi = \left(\frac{\beta_1\left(p_1\frac{\alpha_1}{\beta_1}\right)^{\frac{1}{1-p_1}}}{\beta_2\left(p_2\frac{\alpha_2}{\beta_2}\right)^{\frac{1}{1-p_2}}}\right)^{\frac{(1-p_1)(1-p_2)}{p_2-p_1}} \equiv \chi_F. \tag{4.6}$$

Note that for the equation above to be relevant, we still require $f_1 \leq 1$. This requires either i) $p_2 > p_1$ and $\frac{\beta_1}{\beta_2} > \left[\frac{p_1\alpha_1}{p_2\alpha_2}\right]^{1/p_2}$ or ii) $p_2 < p_1$ and $\frac{\beta_1}{\beta_2} < \left[\frac{p_1\alpha_1}{p_2\alpha_2}\right]^{1/p_2}$.

We note that, within our model, enforcing parity of FPR generally does not minimise the crime rate, i.e. it is not generally true that $\chi_f = 1$ nor $\chi_F = 1$. In fact, for two groups with different p values, these cases can only be achieved for specific relationships between the policing rates of the groups, with the rates generally being required to differ in some ways. This is in contrast to the results of Jung et al., where the policy of minimising crime rate also achieves parity of false positive rates, with equal policing rates. This difference is due to their assumption that the signal distributions for innocents and criminals are the same across groups.

It is interesting to note that it is generally not possible to simultaneously equalise both the false positive rates and false negative rates of the two groups in our model. The false negative rate for group k is

$$FNR_k = 1 - TPR_k = \begin{cases} 1-\alpha_k f_k^{\frac{p_k}{1-p_k}} = 1-\alpha_k\left(p_k\frac{\alpha_k}{\beta_k}\chi_F\right)^{\frac{p_k}{1-p_k}}, & f_k < 1 \\ 1-\alpha_k, & f_k \geq 1. \end{cases} \tag{4.7}$$

In the region where $f_2 \geq 1$, the false negative rates for both groups equalise if and only if their investigation rates for criminals are equal, i.e. $\alpha_1 = \alpha_2$. Assume now that the false positive rates of the two groups

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

are equal. In the region where $f_2 < 1$ but $f_1 \geq 1$, the false negative rates for both groups are equal if and only if

$$\alpha_1 = \alpha_2 \left(p_2 \frac{\alpha_2}{\beta_2} \chi_f \right)^{\frac{p_2}{1-p_2}} \tag{4.8}$$

which is only satisfied when

$$\chi_f = \frac{p_2 \alpha_1}{\beta_2}. \tag{4.9}$$

Recall that χ_f is a constant defined by equation (4.4), so the false negative rates for both groups can be equal only if the parameters happen to satisfy the equation above. Finally, in the region where $f_1 \leq 1$, the false negative rates for both groups are equal if and only if $p_1 = p_2 = p$. When $p_1 = p_2 = p$, the false negative rates are equal for any χ, under the same parameter constraint as the false positive rate being equal. But if $p_1 \neq p_2$, the false negative rates are equal when

$$\alpha_1 \left(p_1 \frac{\alpha_1}{\beta_1} \chi_F \right)^{\frac{p_1}{1-p_1}} = \alpha_2 \left(p_2 \frac{\alpha_2}{\beta_2} \chi_F \right)^{\frac{p_2}{1-p_2}}. \tag{4.10}$$

Simplifying by substituting equation (4.5), the equation above cannot be satisfied.

4.1.1. Existence of solutions

For the remainder of our discussion of solutions with equal false positive rates, we only consider the case $f_1 < 1$, as other solutions require situations in which all individuals are found guilty, which are inherently unsatisfying. Further, we note that some values of χ_F might be unfeasible; that is, a $\chi_F > 1$ may only correspond to solutions that always allow for a lower crime solution for the same τ and κ. In such cases, it is simply not possible for the false positive rates to be matched within the confines of the model. For the sake of argument, assume that χ_F is feasible, generally meaning $\chi_F < 1$. Then we seek a guarantee that there exists a pair (τ, C) that will satisfy equations (4.1) and (4.6) simultaneously. We will now show that the legislator, by appropriately choosing the punishment level κ, can always ensure such a pair, given some conditions specified in the following theorem:

Theorem 4.2. *If (τ_F, C_F) satisfies equation (4.6) and $C_F \leq C_{max}$, where*

$$C_{max} \equiv N_1 \int_0^\infty \Gamma_1(r)\, dr + N_2 \int_0^\infty \Gamma_2(r)\, dr, \tag{4.11}$$

there exists a unique $\kappa = \kappa_F$ such that (τ_F, C_F) satisfies equation (4.1).

Proof. The assumption that (τ_F, C_F) satisfies equation (4.6) gives $f_k = \frac{p_k \alpha_k}{\beta_k} \chi_F$, assumed no larger than 1 for both groups (else equation (4.6) is not relevant). In this case, Δ_k is given solely by the parameters p_k, α_k, and β_k; that is, it is independent of C and τ. Thus, the crime rate given by equation (4.1) under the equalising false positive rate constraint given by (4.6) depends only on κ. For ease of reading, we define $\psi(\kappa)$ to be the RHS of equation (4.1) in this case. Then, since $\Delta_1, \Delta_2 > 0$, $\psi(\kappa)$ monotonically decreases toward zero as κ increases and has a maximum at $\kappa = 0$,

$$C_{max} \equiv \psi(0) = N_1 \int_0^\infty \Gamma_1(r)\, dr + N_2 \int_0^\infty \Gamma_2(r)\, dr. \tag{4.12}$$

Since $\psi(\kappa)$ is continuous, if $C_F \in (0, C_{max}]$, there must exist a unique κ_F such that $\psi(\kappa_F) = C_F$. In other words, (τ_F, C_F) also satisfies equation (4.1), so long as $\kappa = \kappa_F$. □

Corollary 4.3. *Let τ, C, κ satisfy equations (4.1) and (4.6). C monotonically decreases as a function of κ.*

To somewhat reiterate the result of Theorem 4.2, for a given value of χ_F (even if it is not feasible), any desired crime rate $C_F \in (0, C_{max}]$ can be made to result in fair outcomes across groups so long as

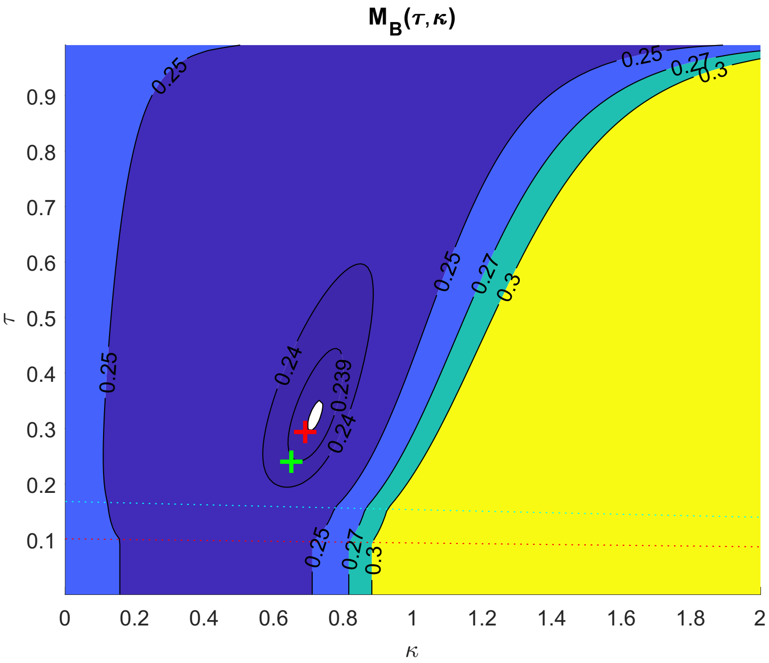


Figure 3. *Contour plot of* $M_B(\kappa,\tau)$ *with* $N_1 = N_2 = 0.5$, $\lambda = 1, n = 9, \alpha_1 = 0.5, \alpha_2 = 0.3, \beta_1 = 0.3, \beta_2 = 0.2, p_1 = 0.2,$ *and* $p_2 = 0.4$. *Here* $\gamma_1, \gamma_2 \sim \mathcal{N}(-2, 3)$. *This figure shows a minimum value of* 0.238 *at* $(\tau,\kappa) = (0.32, 0.72)$. *The green* + *indicates the location of* (κ,τ) *that minimises the mixed objective case. The red* + *indicates the location of* (κ,τ) *that minimises the objective function with the parity of FPR constraint. The red and cyan dotted lines plot the threshold* τ *that makes* $f_1 = 1$ *and* $f_2 = 1$, *respectively.*

τ is chosen to satisfy (4.6) and κ is chosen to satisfy (4.1), independently. Of course, whether this is realisable will depend on whether or not χ_F is feasible. Hence, as seen above in the single group case, if all the judges and legislators desire to do is make the crime rate as small as possible while still being fair, this can be accomplished to any desired level by simply choosing a high enough punishment level and the appropriate threshold. But, as discussed above, arbitrarily increasing punishment levels has the strong downside of leading to arbitrarily large levels of punishment applied to any false positives that might occur. Hence, as before, we will instead consider how the judge and the legislator can minimise the crime rate with some penalty proportional to the false positive rate of each group.

4.2. Judge's and legislator's objective function

Similar to the one-group case, we propose the following two-group objective function:

$$M_B = C + \lambda\kappa^n \left(N_1(1 - C_1)FPR_1 + N_2(1 - C_2)FPR_2\right). \tag{4.13}$$

Analytically describing the potential global minima of M_B is more complicated than the one-group case, as terms containing the various Γ distributions cannot be eliminated during the algebraic manipulation, as they can in the one-group case. However, numerical explorations show in Figure 3 that this objective function can admit a minimum in the interior of the parameter space in at least some scenarios, as seen for a single group. As in the single group case, in this numerical simulation, we first solve the crime rate equation, now for two groups, and then compute the objective function M_B. We choose the parameters, specified in the caption of Figure 3, to ensure that both groups are in the case where not all investigated individuals are assigned guilty as in Theorem 4.1. We chose equal population size and

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

$\gamma_1, \gamma_2 \sim \mathcal{N}(-2, 3)$ – the maximum crime rate is 0.25. We found that $(\tau, \kappa) = (0.32, 0.72)$ minimises M_B with value of 0.238; the crime rate is 0.24 and $FPR_1 = 0.045$ and $FPR_2 = 0.042$. In comparison, with the same parameters, and when $\kappa = 0.72$ is fixed, as in Theorem 4.1, the judge can minimise the crime rate even further to 0.7 with a lower threshold $\tau = 0.22$, but with higher false positive rates for both groups: $FPR_1 = 0.076$ and $FPR_2 = 0.085$.

In the limiting case when $p_1 = p_2 = p$, the objective function is minimised by choosing χ exactly as in the one-group case:

$$\chi = \frac{n - \frac{1}{p}}{n - 1} = \chi_o \tag{4.14}$$

as in (3.19). Similarly, the minimum is well defined and exists when $n > \frac{1}{p}$ and either $\frac{p\alpha_1}{\beta_1} < 1$ or $n < \frac{\alpha_1 - \beta_1}{\alpha_1 p - \beta_1}$. The minimum then happens at (κ_c, χ_o) with κ_c satisfying

$$-N_1\Gamma_1(\kappa_c\Delta_1)f_1^{\frac{1}{1-p}}\beta_1\left[\frac{1}{p\chi_o} - 1\right](1 - \lambda\kappa_c^n FPR_1) - N_2\Gamma_2(\kappa_c\Delta_2)f_2^{\frac{1}{1-p}}\beta_2\left[\frac{1}{p\chi_o} - 1\right](1 - \lambda\kappa_c^n FPR_2)$$
$$+ \lambda n\kappa_c^{n-1}(N_1(1 - C_1)FPR_1 + N_2(1 - C_2)FPR_2) = 0.$$

We will now explore two variants of the objective function, one motivated by an interesting result in the one-group case, the other motivated by minimising disparate impact. Recall that in the single group case, when n is large, the judge's optimal threshold is equivalent to just minimising the crime rate, i.e., choosing $\chi = 1$. Based on this, we define the *mixed objective* case: the judge only cares about minimising crime rate C, and therefore chooses $\chi = 1$, while the legislator picks κ to minimise the resulting M_B when $\chi = 1$. Then one can easily show that M_B has a minimum at some finite κ_M in the mixed objective case. This is because, for $\chi = 1$, $FPR_1 = \beta_1\left(\frac{p_1\alpha_1}{\beta_1}\right)^{\frac{1}{1-p_1}}$ and $FPR_2 = \beta_2\left(\frac{p_2\alpha_2}{\beta_2}\right)^{\frac{1}{1-p_2}}$, constants that do not depend on κ. Moreover, Δ_1 and Δ_2 are likewise constants. One can readily show in this case that $\frac{dM_B}{d\kappa} < 0$ as $\kappa \to 0$ and $\frac{dM_B}{d\kappa} > 0$ as $\kappa \to \infty$, indicating a global minimum at some κ_M.

The objective function is similarly simplified in the case where the false positive rates of the two groups are equalised. We define M_F to be the objective function M_B with parity of false positive rates. We have,

$$M_F = C + \lambda\kappa^n\,(1 - C)\,FPR_F, \tag{4.15}$$

where $FPR_1 = FPR_2 = FPR_F$, which is a constant. Exactly analogously to the mixed objective case, it is easily shown that this fair objective function has a global minimum for some finite value of κ.

It is worth considering how enforcing the fairness constraint affects the optimal solution, and thereby impacts the individuals who are false positives of each group. Consider for now that the judge and legislator have some specific values of λ and n in mind for their objective function M_B, and use them, in conjunction with all the other necessary parameters and distributions, to determine optimal values κ_B and τ_B, equivalently κ_B and χ_B. Then the total negative impact on each group k, in terms of erroneous punishment, is given by

$$N_{B,k} = \kappa_B(1 - C_{B,k})FPR_{B,k}. \tag{4.16}$$

Now imagine an alternative scenario in which the same values of λ and n are chosen, but equality of false positives is enforced. The optimal parameters are then given by κ_F and χ_F, where χ_F is given in (4.6). The total negative impact on each group in this case is

$$N_{F,k} = \kappa_F(1 - C_{F,k})FPR_F. \tag{4.17}$$

Then it is natural to ask whether implementing the fairness constraint has increased or decreased the harm to each group; i.e., what is the relationship between $N_{B,k}$ and $N_{F,k}$ for each group k?

As a first point of consideration, note that if $\chi_B < \chi_F$, implementing the fairness constraint will cause the false positive rates of both groups to increase, with one increasing more than the other to make them

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

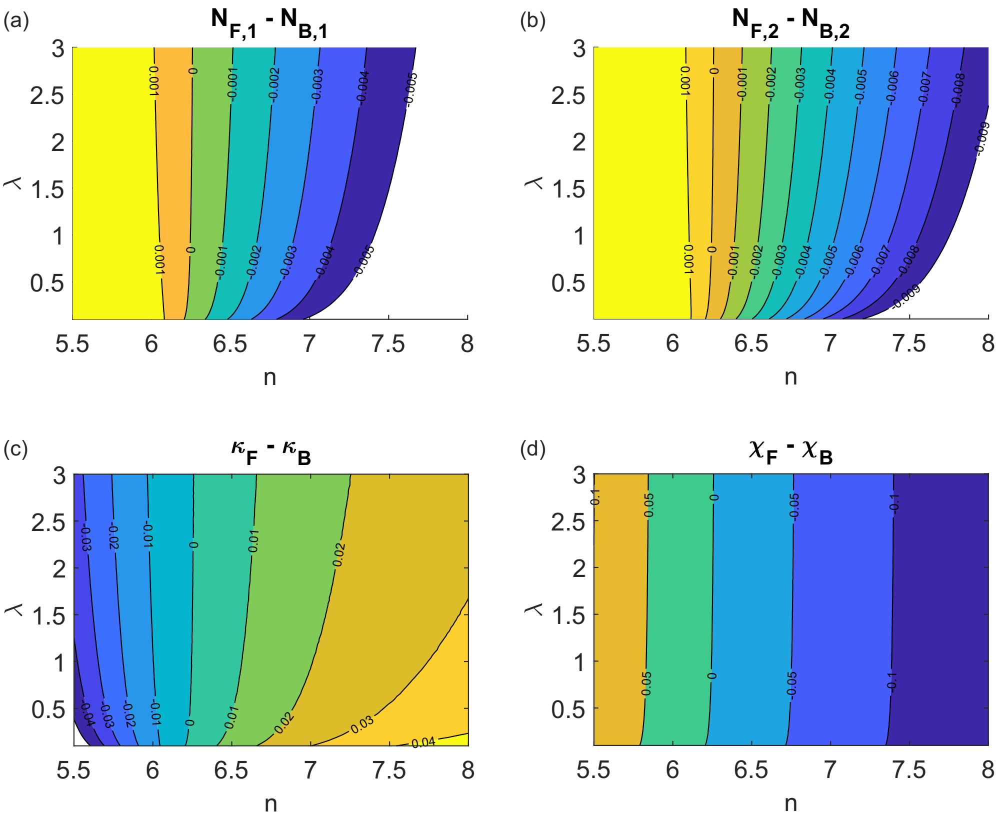


Figure 4. Exploring the impact of the fairness constraint. (a) and (b) plot the difference in negative impact of satisfying the unconstrained objective function and the fairness-constrained objective function – (c) and (d) plot the difference between the justice system's optimal choice. Here, $N_1 = N_2 = 0.5, \alpha_1 = 0.5, \alpha_2 = 0.3, \beta_1 = \beta_2 = 0.1, p_1 = 0.2, p_2 = 0.4$ *and* $\gamma_1, \gamma_2 \sim \mathcal{N}(-2, 3)$*; these parameters give* $\chi_F = 0.48$*.*

equal. On the other hand, if $\chi_B > \chi_F$ then implementing the fairness constraint causes both false positive rates to decrease. This of course does not paint a complete picture, as the behaviour of κ_B and κ_F will greatly affect the negative impact. We therefore turn to numerical simulations to gain insight.

We show in Figure 4 the results of a numerical study where, for fixed group parameters, we have computed $N_{B,k}$ and $N_{F,k}$ and created a contour plot of the difference between them for varying values of n and λ. Interestingly, for certain regions in this plot, both groups benefit (in terms of negative impact) by the implementation of the fairness constraint. Conversely, in the remaining region both groups are harmed by implementing the fairness constraint. Based on this, it seems clear that a fairness constraint should only be considered for certain combinations of λ and n where both groups benefit from its implementation, and should certainly not be considered outside of these.

On the other hand, if there is no a priori reason to choose any very specific values for λ and n, but a fairness constraint is desired, then Figure 4 also shows that there is a curve in parameter space along which the optimal solutions to the unconstrained case and the constrained case are identical. That is, if n and λ are chosen from that curve, then the fairness constraint is a natural side effect of the unconstrained optimisation problem. Of course, whether or not such a curve will exist for other group-dependent parameter values is not necessarily guaranteed, and indeed in cases where χ_F is not feasible it cannot exist.

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

5. Discussions

In this work, we explored a model where rational agents choose whether or not to commit a crime while the justice system attempts to influence this choice. In our model, the justice system's attempts are carried out by the judge setting a single society-wide guilt threshold, influencing the likelihood of conviction and the legislator setting a society-wide punishment level, both of which are applied after calculating the probability of guilt based on an individual's signal and the crime rate. For the case of a single societal group, when the signals of the innocents and the criminals are more distinguishable than their respective policing rates, the judge is able to minimise the crime rate by choosing a threshold in which some people are not found guilty. In other words, the judge does not need to set the harshest threshold in order to minimise the crime rate. However, we showed that given this minimum crime rate, the legislator can further decrease the crime rate solely by increasing the level of punishment – the legislator will punish everyone as harshly as possible if minimising crime is the only concern. To avoid such a scenario, we proposed that both the judge and the legislator minimise an objective function containing both the crime rate and a quantity proportional to the amount of punishment given to innocents. In doing so, we find that, depending on the parameters chosen in this objective function, there can exist a unique optimal evidence threshold and punishment level. But, different societies may choose different values for the parameters of the objective, explaining in part why punishment levels and evidence thresholds can vary from society to society, even if the societies seem similar in other respects.

We then explored the case in which there are two distinct societal groups with different signal distributions and/or policing rates. The two-group case allows us to study a fairness notion of equalising false positive rates across the two groups, thus extending the theoretical framework of criminal deterrence where fairness notions are considered, specifically the case of the *disparate impact* fairness notion. We found that, under some group-dependent parameter combinations, achieving this notion of fairness is impossible. However, in other cases, the judge and legislator can achieve this goal through careful choice of threshold and punishment level. We showed that, similar to the single group case, solely minimising crime rate leads to draconian punishment by the legislators. We then proposed an objective function similar to the single group case and showed through a numerical simulation that for some parameter combinations, the judge chooses a conviction threshold which still does not find everyone guilty and the legislator chooses a punishment level that is not the harshest and also not the most lenient. Finally, we showed that imposing the fairness constraint within the optimisation problem would generally lead to outcomes that either benefit or harm both groups, depending on the choice of the objective function's parameters, but that there exist parameter combinations where the constrained and unconstrained problems give the same optimal choices. This finding is perhaps the most sociologically important within our work, as it indicates that simply imposing certain kinds of fairness between groups, while morally appealing, might in fact have the unintended consequence of harming both groups, relative to the case where fairness is not enforced. That said, the finding also indicates that, at least in some cases, the criminal-justice objective function can be tailored such that this does not occur, and such that minimising the objective automatically leads to this kind of fairness. Such tailoring may lead to a variety of different optimal behaviours, though, some of which might be considered unappealing otherwise (e.g., low thresholds for guilt or very severe punishment levels).

The work we present is a toy model of a very complicated system, with many simplifying assumptions made for the sake of tractability, any of which could potentially be relaxed and explored in future work. One such assumption, mentioned previously, is that the legislator chooses one punishment level κ, when in practice, the legislator often chooses a range of punishment levels as a guide for the judge. Within the context of our model, this could naturally arise if there is disagreement among legislators or judges on what the parameters λ and n should be within the objective function. Then, no single optimal κ would exist. As a consequence, individuals in the society would need to factor into their decision whether or not to commit crime the various punishments they might meet, along with the probability of each. Moreover, we assume that individuals here act fully rationally and have perfect information of the conviction threshold and punishment level to calculate their utility function. In reality, however,

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

there are other considerations aside from utilities in choosing to commit or not to commit crime. One example of a perhaps powerful influence is local social norms. Further, it is practically impossible for a criminal to calculate their expected utility from committing or not committing a crime; there are other decision-making models that may be able to incorporate this uncertainty [24–8]. In our model, we do not distinguish between the severity of crimes. It might be interesting to have an independent way for the severity of the crime to both influence an individual's utility function and the punishment incurred if convicted. We considered the judge and legislator to be optimising an objective function that includes only the crime rate and a penalty term for false punishments. In reality, there might be other factors the judge and legislator need to consider such as cost of conviction and cost of punishment, among others – as discussed in [1, 22]. Lastly, our model shows how the likelihood of conviction and level of punishment deter crime. But there are other ways to approach the crime problem. For example, legislators can create laws to promote social welfare and create more options for individuals to get rewards from non-criminal activities, known in the literature as positive reinforcements – something that has been explored in the criminal deterrence literature [5, 2–15], although not as emphasised according to [22]. Incorporating any of these ideas could lead to interesting future work.

Competing interests. The authors declare none.

Financial support. None.

References

[1] Becker, G. S. (1968) Crime and punishment: An economic approach. *J. Political Econ.* **76**(2), 169–217.
[2] Berenji, B., Chou, T. & D'Orsogna, M. R. (2014) Recidivism and rehabilitation of criminal offenders: A carrot and stick evolutionary game, ISSN. *PLoS ONE* **9**(1), e85531. DOI: 10.1371/journal.pone.0085531. ISSN 1932-620.
[3] Chalfin, A. & McCrary, J. (2017) Criminal deterrence: A review of the literature. *J. Econ. Lit.* **55**(1), 5-0022–0048-0515. DOI: 10.1257/jel.20141147.
[4] Chouldechova, A. (2017) Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* **5**(2), 153–163. DOI: 10.1089/big.2016.0047. ISSN 2167-647X.
[5] Curry, P. A. & Doyle, M. (2016) Integrating market alternatives into the economic theory of optimal deterrence. *Econ. Inq.* **54**(4), 1873–1883. DOI: 10.1111/ecin.12344. ISSN 1465-729.
[6] Fryer, R. G. (2019) An empirical analysis of racial differences in police use of force. *J. Polit. Econ.* **127**(3), 1210–1261. DOI: 10.1086/701423. ISSN 1537-534X.
[7] Hu, L. (2025) Does calibration mean what they say it means; or, the reference class problem rises again. *Philos. Stud.*, 1–27.
[8] (2002). *Game Theory and Decision Theory in Agent-Based Systems*, Springer US. DOI: 10.1007/978-1-4615-1107-6. ISBN 9781461511076.
[9] Jung, C., Kannan, S., Lee, C., Pai, M., Roth, A. & Vohra, R. (2020) Fair prediction with endogenous behavior. In: *Proceedings of the 21st ACM Conference On Economics and Computation, EC '20*. DOI: 10.1145/3391403.3399473.
[10] Kleiman, M. & Kilmer, B. (2009) The dynamics of deterrence. *Proc. Natl. Acad. Sci.* **106**(34), 14230–14235. DOI: 10.1073/pnas.0905513106. ISSN 1091-6490.
[11] Kleinberg, J., Mullainathan, S. & Raghavan, M. (2017) Inherent trade-offs in the fair determination of risk scores. *Schloss Dagstuhl – Leibniz-Zentrum für Informatik*, DOI: 10.4230/LIPICS.ITCS.2017.43.
[12] Knowles, J., Persico, N. & Todd, P. (2001) Racial bias in motor vehicle searches: Theory and evidence. *J. Polit. Econ.* **109**(1), 203–229. DOI: 10.1086/318603. ISSN 1537-534X.
[13] Knox, D., Lowe, W. & Mummolo, J. (2020) Administrative records mask racially biased policing. *Am. Polit. Sci. Rev.* **114**(3), 619–637. DOI: 10.1017/s0003055420000039. ISSN 1537-5943.
[14] Loi, M. & Heitz, C. (2022) Is calibration a fairness requirement?: An argument from the point of view of moral philosophy and decision theory. In: *2022 ACM Conference On Fairness, Accountability, and Transparency, FAccT '22*, pp. 2026–2034. DOI: 10.1145/3531146.3533245.
[15] Mungan, M. C. (2019) Positive sanctions versus imprisonment. *SSRN Electron. J.* DOI: 10.2139/ssrn.3317552. ISSN 1556-5068.
[16] Persico, N. (2002) Racial profiling, fairness, and effectiveness of policing. *Am. Econ. Rev.* **92**(5), 1472–1497. DOI: 10.1257/000282802762024593. ISSN 0002-8282.
[17] Petersilia, J. (1985) Racial disparities in the criminal justice system: A summary. *Crime Delinquency* **31**(1), 15–34. DOI: 10.1177/0011128785031001002. ISSN 1552-387X.
[18] Polinsky, A. M. (2015) Deterrence and the optimality of rewarding prisoners for good behavior. *Int. Rev. Law Econ.* **44**, 1–7. DOI: 10.1016/j.irle.2015.04.004. ISSN 0144-8188.
[19] Polinsky, A. M. & Shavell, S. (1984) The optimal use of fines and imprisonment. *J. Public Econ.* **24**(1), 89–99 DOI: 10.1016/0047-2727(84)90006-9. ISSN 0047-2727.

https://doi.org/10.1017/S0956792525100296 Published online by Cambridge University Press

[20] Polinsky, A. M. & Shavell, S. (2000) The economic theory of public enforcement of law. *J. Econ. Lit.* **38**(1), 45–76 DOI: 10.1257/jel.38.1.45. ISSN 0022-0515.

[21] Raskolnikov, A. (2013) Irredeemably inefficient acts: A threat to markets, firms, and the fisc. *Georget. Law J.* **102**, 1133.

[22] Raskolnikov, A. (2020) Criminal deterrence: A review of the missing literature. *Supreme Court Economic Review* **28**, 1–59. DOI: 10.1086/710158. ISSN 2156-6208.

[23] Rizzolli, M. & Stanca, L. (2009) Judicial errors and crime deterrence: Theory and experimental evidence. *SSRN Electron. J.* DOI: 10.2139/ssrn.1441325. ISSN 1556-5068.

[24] Rossmo, D. K. & Summers, L. (2022) Uncertainty and heuristics in offender decision-making: Deviations from rational choice. *J. Crim. Just.* **81**, 101923–2352. DOI: 10.1016/j.jcrimjus.2022.101923. ISSN 0047-2352.

[25] Simoiu, C., Corbett-Davies, S. & Goel, S. (2017) The problem of infra-marginality in outcome tests for discrimination, ISSN. *Ann. Appl. Stat.* **11**(3). DOI: 10.1214/17-aoas1058. ISSN 1932-6157

[26] Stigler, G. J. (1970) The optimum enforcement of laws. *J. Polit. Econ.* **78**(3), 526–536. http://www.jstor.org/stable/1829647. ISSN 00223808.

[27] Thorstad, D. (2024) *Inquiry Under Bounds*, Oxford University PressOxford. DOI: 10.1093/oso/9780198886143.001.0001. ISBN. 9780191994227.

[28] Vul, E., Goodman, N., Griffiths, T. L. & Tenenbaum, J. B. (2014) One and done? Optimal decisions from very few samples. *Cogn. Sci.* **38**(4), 599–637. DOI: 10.1111/cogs.12101. ISSN 1551-6709.

Cite this article: Nurhan Y. and Short M.B. Modelling crime response to deterrence: Existence of solutions, optimal policies, and fairness. *European Journal of Applied Mathematics*, https://doi.org/10.1017/S0956792525100296