Crime, Punishment, and Prejudice[*]

Philip A. Curry[†]
Simon Fraser University

Tilman Klumpp[‡]
Emory University

June 2008

Abstract

We examine the link between the penalties used to punish convicted criminals, and judicial prejudice against defendants. In our model, agents choose to commit crimes if their privately observed utility from doing so is high enough. A crime generates noisy evidence, and defendants are convicted when the realized amount of evidence is sufficiently strong to establish the probability of guilt beyond a fixed threshold. We show that if convicted offenders are incarcerated, poorer individuals face a strong prior prejudice in trials and are therefore convicted with less evidence than members of the other group. At the same time, they commit crimes more frequently. Penalties such as monetary fines can eliminate this bias, but may also reverse it. We fully characterize the penalty schedule that guarantees an unbiased equilbrium. We extend the model such that agents also differ in characteristics such as race or gender. We show biased outcomes (targeted at subgroups of the population) may still exist, even if all members of the population are ex-ante alike in their economic characteristics.

Keywords: Crime, punishment, statistical discrimination, stereotypes, prejudice.

JEL code: D82, J71, K14, K42.

"It is in justice that the ordering of society is centered."

— Aristotle

1   Introduction

Criminal conviction rates in the United States differ drastically across racial groups. The jail incarceration rates for African-Americans, for example, is 800 people per 100,000, while the rate for white Americans is 166 per 100,000. An estimated 12% of U.S. black males in their late twenties were incarcerated in 2005, as opposed to 1.7% of white males.[1] That these differences exist is undisputed. It is much less clear, however, *why* they exist. One possible explanation is that blacks commit more crimes than whites. For example, criminal participation tends to be correlated with economic characteristics such as income, education, or area of residence, and these characteristics differ across racial groups. Another possibility is that the criminal justice system is somehow "biased" in that blacks are more easily convicted than whites are.

This paper pursues a twofold objective. First, we demonstrate that differences in crime rates as well as judicial prejudice against certain individuals can arise simultaneously in a single model. Second, we ask whether our model can inform policies aimed at reducing racial differences in the criminal justice system. The suggestions we derive from our framework are novel and perhaps unexpected: To understand, and perhaps reduce, the aforementioned differences, one should consider the way by which certain crimes are punished.

In our model, a Bayesian jury must decide whether a given amount of evidence is sufficient to convict a defendant. Suppose that the defendant is sent to prison if convicted. Because the deterrence effect of incarceration increases in the defendant's wealth, richer individuals are ex-ante less likely to commit crimes than poorer ones. Economic variables such as the defendant's income or social status may hence enter the jury's prior belief (i.e. their prejudice) about the guilt of the defendant.[2] Our framework thus exposes a new mechanism by which economic differences across individuals can cause differences not only in criminal behavior, but also in how defendants are treated in the courts. We then consider the possibility that income differences not only generate prejudice, but are also the result of such prejudice. For example, persons believed to be more prone to criminal behavior may have a harder time finding well-paying employment opportunities. Such persons are therefore poorer and less likely to be deterred by the threat of losing their income, justifying the initial prejudice. If there are multiple equilibria, in each of which an individual's income is consistent with the

---

[1]Source: Bureau of Justice Statistics Prison and Jail Inmates at Midyear 2005.

[2]We assume that convictions are issued if only if the *posterior* probability of guilt is beyond a certain threshold of reasonable doubt, which is the same for all individuals. However, since there are differences in *prior* beliefs, the amount of evidence required to convict a defendant must differ across defendants.

judicial bias against the individual, a social norm may develop that associates different subgroups of the population with different such equilibria. In such an overall "stereotyping equilibrium," race or ethnicity serves as a coordination device through which individual roles and commonly expected behaviors are developed.[3] Corrective policies aimed solely at the labor market, such as affirmative action, may then have a limited effect because of discrimination in the judicial system.

We show that other forms of punishment can, in principle, eliminate such outcomes. In particular, the adoption of a schedule of monetary fines (which may depend on the defendant's wealth) as a means of punishment may reduce unfavorable prejudice. For certain felonies such monetary penalties may indeed be considered an alternative to incarceration.[4] We remark that, at this point, we do not undertake any formal welfare analysis of different approaches to punishing criminals. We do not formally consider the costs to either crime or corrections. There are many reasons, besides the one examined in this paper, why one form of punishment may be preferable over another. Our concern here is with which types of penalties can lead to judicial biases, and which ones do not.

Our paper is related to several strands of theoretical literature (empirical papers, as they relate to our assumptions and results, will be discussed later in Section 3 and Section 4.) Our work contributes to the research on the optimal use of fines and imprisonment (e.g. Becker, 1968; Posner, 1992; Polinsky and Shavell, 1984; Morris and Tonry, 1990; Levitt, 1997; and many others). No attempt of a thorough review of this extensive literature will be made here. The idea that the deterrence effect of a jail sentence may depend on the defendant's income appears in Lott (1987), and Spiegel and Templeman (1989); however, neither considers how this affects jury beliefs in equilibrium, or how an unbiased penalty schedule can be constructed. Our stereotyping argument is related to the literature on statistical discrimination in labor markets, which includes seminal contributions by Phelps (1972) and Arrow (1973), and important applications by Coate and Loury (1993) and others. An application to crime is Verdier and Zenou (2004), who examine a model of statistical discrimination, location choice, and criminal activity; however, their paper neither examines Bayesian updating in the courts nor the impact of various forms of penalties on beliefs. Finally, several papers consider racial profiling by the police (e.g. Knowles et al., 2001; Persico, 2002; Alexeev and Leitzel, 2004; Bjerk, 2007). These papers assume ex-ante differences in criminal behavior across racial subgroups and are not concerned with how such difference might come to be.[5]

---

[3]In economic terms, racial labels are "sunspots": economically meaningless events, which derive significance only because they coordinate individual actions and social expectations.

[4]For example, West Germany in 1969 eliminated prison sentences of under six months from its penal code and, except for rare circumstances, replaced them with day fines. Several other European countries, and some Latin American countries, have adopted similar systems, and U.S. courts have been experimenting with replacing jail time by montary sanctions since the 1980s (Hillsman, 1990).

[5]There is also a literature on *false* beliefs by prosecutors and judges (e.g. Georgakopoulos, 2004; Burke, 2007). In our paper, on the other hand, all beliefs will be correct in equilibrium.

We proceed as follows. Section 2 describes our theoretical model, assuming that agents' economic characteristics are exogenously given and observable. In Section 3 we derive a number of results for this case, as well as various comparative results with respect to different forms of penalties. In Section 4 we look at the possibility of statistical discrimination when there is a feedback from prejudice to income. In Section 5, we discuss some policy implications of our model. In particular, we contrast our findings with possible interventions in the labor market or in the courts directly. Most proofs are in the Appendix.

2   A Model of Crime, Punishment, and Prejudice

In our model, an individual must decide whether to commit a crime, and a judge or jury must decide whether to convict and punish a defendant accused of committing a crime. The individual is characterized by an exogenously given and publicly observable variable $w \in [\underline{w}, \infty)$ (where $\underline{w} > 0$), which we refer to as the individual's type. Ceteris paribus, individuals of a higher type receive a higher payoff, so that one may regard w as an agent's wealth, income, or social status. The events unfold as follows.

1. The individual comes across an opportunity to commit a crime. The benefit to committing this crime is $\eta \in [0, \infty)$, which is privately observable and drawn according to distribution Q. Q is continuous on $[0, \infty)$ and independent of w. The individual's decision is denoted $d \in \{0, 1\}$. If the crime is committed ($d = 1$) the individual receives the benefit η.

2. If the individual has committed a crime, an investigation is initiated with certainty. If the individual has not committed a crime, there is a probability $\lambda \in (0, 1)$ that an investigation is initiated "by accident." The assumption of an accidental investigation introduces the possibility that a person who faces trial is innocent.[6]

3. If under investigation, a random amount of evidence $t \in [0, 1]$ against the individual is discovered. If the individual has committed the crime, t is a random draw from distribution F. In case of accidental investigation, t is drawn from G. We assume that F and G have support $[0, 1]$, are continuous with densities f and g, and that $f(t)/g(t)$ increases strictly in t. A higher value of t hence means stronger

---

[6]If this possibility did not exist, every defendant would be guilty and criminal trials would be meaningless. There are several possibilities why innocent persons may be investigated for a crime. First, a person may be charged with a crime that has in fact happened, but which was committed by somebody else. Second, an individual may be under investigation for events which may or may not be the result of a crime. Consider, for example, an area such as tort law: Damaging events may take place which may or may not be the result of negligence, and at trial it must be determined whether the individual should be held responsible. Thus, even if the individual did in fact behave responsibly, it is possible that he is tried and held responsible for damages that were caused by chance.

evidence against the individual. We further make the technical assumptions that $0 < f(0) < \infty$ and $0 < g(0) < \infty$.

4. After the evidence is discovered, the individual becomes a defendant and has to stand trial. A jury observes w as well as t and forms belief $\theta(w, t) = P[d = 1|w, t]$, which is the probability that the individual is guilty conditional on observables. The individual is convicted if $\theta(w, t) \geq \alpha$, where $\alpha \in (0, 1)$ represents the standard of proof.[7]

If the individual is not investigated, or if he is investigated and subsequently acquitted, he receives utility $u(w)$. We assume that u is twice differentiable with $u'(w) > 0$ and $u''(w) \leq 0$.

2.1 Penalties

If the jury convicts the defendant (rightfully or wrongfully), the defendant is sentenced to a punishment which reduces the payoff to the individual. We model punishments by means of a *penalty schedule*, which is a mapping

$$\rho : [\underline{w}, \infty) \to \mathbb{R}$$

such that $\rho(w) \in [0, w]$. The interpretation is that, when sentenced, the individual's type is decreased by the amount $\rho(w)$. By using the concept of a penalty schedule, we can examine many important cases of penalties in our model:

Example 1. To characterize a *(simple) fine*, let δ be the constant amount an individual must pay if convicted. Then the fine corresponds to the penalty schedule $\rho(w) = \delta$. For all individuals to be able to pay this fine, we assume that $\delta \leq \underline{w}$.

Example 2. To characterize a *prison sentence*, let u_0 be the fixed utility an individual receives in prison; and set $w_0 = u^{-1}(u_0)$. Then a prison sentence corresponds to the penalty schedule $\rho(w) = w - w_0$. We assume that $u_0 > -\infty$.

Example 3. More generally, ρ can be a means-adjusted penalty. For example, a *proportional fine* is given by the penalty schedule $\rho(w) = \gamma w$, where $\gamma \in (0, 1)$ is the fraction of income taken away from convicted individuals.

2.2 The jury

At trial, the jury computes a belief of the defendant's guilt through Bayesian updating. Specifically, let $p(w)$ denote the jury's prior belief that an individual of income w commits a crime. We will call $p(w)$ the *prejudice* held against individuals of income w.

---

[7]The interpretation is that courts must determine whether or not the evidence establishes the defendant's guilt beyond a "reasonable doubt." In our model, α quantifies what reasonable means.

The posterior Bayesian likelihood that the investigated individual is guilty, conditioning on evidence t, is then

$$\theta(w,t) \equiv P[d=1|w,t] = \frac{p(w)f(t)}{p(w)f(t) + \lambda(1-p(w))g(t)} \in [0,1]. \tag{1}$$

Given that $f(t)/g(t)$ increases, $\theta(w,t)$ increases in t and in $p(w)$. Now recall that we assume that the jury convicts the defendant if and only if the evidence results in a posterior likelihood of the defendant's guilt of at least α. Given α, let $t(w)$ be such that

$$\alpha = \theta(w,t(w));$$

$t(w)$ is then the *conviction threshold* that is applied to individuals of income w. Note that α is the same for all defendants. However, because the jury's prior belief depends on the defendant's type w, the actual amount of evidence $t(w)$ required to prove guilt beyond probability α depends on w.

2.3 The individual's decision

Let $m_1(w)$ denote the probability that an individual of income w is convicted conditional on having committed the crime, and let $m_0(w)$ denote the probability that the same individual is (wrongfully) convicted conditional on not having committed the act. We can express the probability of conviction following a crime from the individual's perspective as

$$m_1(w) = P[\theta(w,t) \geq \alpha|d=1] = P[t \geq t(w)|d=1] = 1 - F(t(w)),$$

and the probability of wrongful conviction as

$$m_0(w) = \lambda P[\theta(w,t) \geq \alpha|d=0] = \lambda P[t \geq t(w)|d=0] = \lambda(1 - G(t(w))).$$

The crime is then committed if and only if

$$\eta > q(w) \equiv [m_1(w) - m_0(w)]\, \Delta(w),$$

where

$$\Delta(w) = u(w) - u(w - \rho(w))$$

denotes the utility loss the individual experiences when sentenced under the penalty schedule ρ. The variable $q(w)$ is the expected cost of committing the crime, so that the defendant decides to commit the act when the benefit of doing so, η, exceeds the expected cost, $q(w)$. We call $q(w)$ the *decision threshold* for the individual.

2.4 Rational expectations equilibrium

An equilibrium of this model consists of three elements: A prior belief, or prejudice, for the jury $(p(w))$, a decision threshold for the individual $(q(w))$, and a conviction threshold for the jury $(t(w))$. Note that these elements are functions of the individual's type w. We say that the tuple (p^*, q^*, t^*) constitutes a *rational expectations equilibrium* if it satisfies the following conditions:

1. The prejudice toward defendants with income w is consistent with the probability that these agents actually commit crimes:

$$p^*(w) = 1 - Q(q^*(w)) \quad \forall w. \tag{2}$$

2. The decisions of agents with income w to commit crimes are optimal, given the conviction thresholds $t^*(w)$ applied to these individuals:

$$q^*(w) = [1 - F(t^*(w)) - \lambda(1 - G(t^*(w)))] \Delta(w) \quad \forall w. \tag{3}$$

3. The jury convicts a defendant if and only if the evidence establishes the defendant's guilt beyond probability α, where this probability is computed by Bayes' Rule using the prejudice $p^*(w)$ as the prior:

$$\theta(w, t^*(w)) = \alpha \quad \forall w. \tag{4}$$

3 Equilibrium Analysis

The following result establishes existence of equilibrium, as well as a condition for uniqueness.

Lemma 1. *A rational expectations equilibrium exists. Furthermore, if $\lambda \leq \overline{\lambda} \equiv f(0)/g(0)$, the equilibrium is unique.*

We now examine how the equilibrium values $p^*(w)$, $q^*(w)$, and $t^*(w)$ depend on the individual's type w. Throughout, we assume that the probability of accidental investigation is small enough $(0 < \lambda \leq \overline{\lambda}$) so that a unique equilibrium is guaranteed by Lemma 1.[8] We say that the equilibrium is *biased against lower types* if $w > w'$ implies

$$p^*(w) < p^*(w'), \quad t^*(w) > t^*(w'), \quad q^*(w) > q^*(w').$$

---

[8] If the probability of an accidental investigation were large enough, then there could exist an equilibrium in which a large number of people believe they will be convicted even if they don't commit the crime, so the crime rate is high and jurors require little evidence to convict, supporting the original beliefs. This equilibrium would coexist with another one in which people did not believe there is much chance that they would be convicted if they do not commit the crime, and so jurors require a lot of evidence, again justifying the original beliefs.

That is, the jury favors richer defendants and applies a lower conviction threshold to poorer ones. Conversely, richer defendants are less likely to commit a crime. An equilibrium is *biased against higher types* if the reverse inequalities hold. Finally, if p^*, q^* and t^* are constant, the equilibrium is *unbiased*.

3.1 Penalties and equilibrium bias

An individual's type w enters the equilibrium in condition (3), through the utility loss $\Delta(w)$. If Δ increases—for example, if prison is used as punishment—richer defendants are penalized more severely than poorer ones if convicted. One should then expect that these individuals are more deterred by the threat of punishment and hence commit crimes less frequently. The opposite should be expected if Δ was decreasing in w. While this argument is quite intuitive, it is important to recognize that we are looking at an *equilibrium* relationship between several variables. In particular, the utility loss effect is fully taken into account at trial: Jurors who realize that some individuals are more severely affected by a particular punishment than others will think it less likely that these individuals commit a crime. In equilibrium, therefore, jurors will entertain a prejudice that favors those who experience large losses from punishment. These types are less likely to be convicted should they commit a crime, which acts against the larger utility loss if convicted. This effect creates a countervailing incentive to commit *more* crimes. A priori, it is not obvious which of the two effects is stronger. Our next result shows that the net effect is in the same direction as the utility loss effect:

Lemma 2. *The equilibrium is (a) biased against higher types if $\Delta(w)$ strictly decreases in w, (b) biased against lower types if $\Delta(w)$ strictly increases in w, and is (c) unbiased if $\Delta(w)$ is constant.*

In order to determine whether a bias can arise for a certain penalty schedule, we need to examine the monotonicity properties of Δ implied by the penalty. To achieve an unbiased equilibrium, the penalty schedule ρ must be such that Δ is constant: $\Delta(w) = \overline{\Delta}$ for all w, where $\overline{\Delta}$ is the utility loss inflicted on every convicted defendant regardless of his type w. Differentiating Δ, we obtain

$$\Delta'(w) = u'(w) - (1 - \rho'(w))u'(w - \rho(w)). \tag{5}$$

For unbiasedness we need $\Delta'(w) = 0$, so that the penalty schedule ρ must satisfy the differential equation

$$\rho'(w) = \frac{u'(w - \rho(w)) - u'(w)}{u'(w - \rho(w))}. \tag{6}$$

To solve for ρ, one requires an initial condition. This can be obtained by computing the penalty $\rho(\underline{w})$ required to achieve the desired utility loss for the lowest type: $\Delta(\underline{w}) = u(\underline{w}) - u(\underline{w} - \rho(\underline{w})) = \overline{\Delta}$, or

$$\rho(\underline{w}) = \underline{w} - u^{-1}(u(\underline{w}) - \overline{\Delta}). \tag{7}$$

The differential equation (6) describes how to trace out the penalty schedule ρ that maintains the same utility loss $\overline{\Delta}$ for all types above $\underline{w}$.[9] This condition represents a knife-edge case, in that it characterizes exactly those penalty schedules which lead to unbiased equilibria. Departing from the knife-edge case in either direction leads to biased equilibria. Specifically, using the same steps as above, the equilibrium is biased against lower types if

$$\rho'(w) > \frac{u'(w - \rho(w)) - u'(w)}{u'(w - \rho(w))}. \tag{8}$$

and biased against higher types if

$$\rho'(w) < \frac{u'(w - \rho(w)) - u'(w)}{u'(w - \rho(w))}. \tag{9}$$

From (6)–(9) it is apparent that the direction of equilibrium bias depends on the utility function u as well as the penalty schedule ρ. Of course, if the utility loss Δ induced by u and ρ is non-monotonic in w, the equilibrium is neither biased nor unbiased according to our definitions. In this case, there would be a local bias against lower types at some values of w, and a local bias against higher types at other values of w.

3.2   Prison vs. fines

As shown earlier, our model encompasses many types of punishments commonly used in the real world, such as prison sentences, simple fines, and means-adjusted fines. In this section, we examine if these types of punishments lead to biased equilibria. The next result states a number of sufficient conditions for unbiased and biased equilibria.

Theorem 3. *The equilibrium is*

   (a) unbiased if punishment is by a simple fine and individuals are risk-neutral,

   (b) unbiased if punishment is by a proportional fine and individuals have logarithmic utility,

   (c) biased against lower types if punishment is by imprisonment, and

   (d) biased against higher types if punishment is by a simple fine and individuals are strictly risk averse (i.e. $u''(w) < 0 \; \forall w$).

---

[9]Note that (6)–(7) describe the solution to a mechanism design problem which is quite similar to, say, the problem of designing an auction or other allocation mechanism: Similar to an incentive compatibility constraint, the unbiasedness requirement leads to a solution in terms of the slope of the design object (the penalty schedule in this case). Similar to the individual rationality constraint, the fact that a certain deterrence effect must be created for the lowest type yields an initial condition.

Thus, imprisonment automatically leads to a bias against lower types, and the opposite is true for simple fines if agents are risk averse. A proportional fine, on the other hand, is an intermediate form of punishment, and Theorem 3 (b) identifies a special case for which it leads to unbiased equilibria. Note that with logarithmic utility, the utility loss from a proportional fine can be written as

$$\Delta(w) = \ln w - \ln(1-\gamma)w = \ln w - \ln(1-\gamma) - \ln w = -\ln(1-\gamma),$$

which is independent of w. Generally, however, a proportional fine can result in an equilibrium bias.

To examine the effects of different forms of punishment further, we now derive a simple sufficient condition for equilibrium bias in terms of risk preferences and elasticities. First, let $R(w)$ denote the coefficient of relative risk aversion of u at w:

$$R(w) = -w\frac{u''(w)}{u'(w)}.$$

Observe that risk-neutral utility has $R(w) = 0$, and logarithmic utility has $R(w) = 1$. Second, let $\varepsilon(w)$ denote the elasticity of ρ at w:

$$\varepsilon(w) = w\frac{\rho'(w)}{\rho(w)}.$$

Observe that a simple fine has $\varepsilon(w) = 0$, a proportional fine has $\varepsilon(w) = 1$, and a prison sentence has $\varepsilon(w) > 1$ if $w_0 > 0$. We have the following result:

Theorem 4. *The equilibrium is (a) biased against higher types if $R(w) > 1$ and $\varepsilon(w) \leq 1 \ \forall w$, and (b) biased against lower types if $R(w) < 1$ and $\varepsilon(w) \geq 1 \ \forall w$.*

Note that the "separating line" between cases (a) and (b) in Theorem 4 is the knife-edge case of logarithmic utility and a proportional fine. Suppose now that *relative* risk aversion increases above the logarithmic case. In this case, poorer individuals will experience more disutility if the same fraction of their income is confiscated, compared to richer individuals. Similarly, if we assume logarithmic utility but decrease the income elasticity of the penalty below the proportional case, poorer individuals are again hurt more by the penalty than richer ones. Thus, when both changes occur together, the deterrence effect Δ unambiguously decreases in w, and as Theorem 4 (a) states the equilibrium will be biased against higher types. An analogous intuition applies to case (b), of course.

Table 1 summarizes our results in the previous two Theorems. The superscripts indicate the relevant theorems. (If the utility function u is such that its degree of risk aversion is not uniformly above or below one, then Table 1 is still informative about the *local* bias in a neighborhood of a certain value of w.)

The practical relevance of our results hinges on whether the various forms of punishments do in fact impart a differential deterrence effect on individuals with differing

9

Preferences	Penalties		
	Simple Fine	Proportional Fine	Imprisonment
$R(w) = 0$	unbiased [3(a)]	against low [4(b)]	against low [3(c)]
$R(w) \in (0, 1)$	against high [3(d)]	against low [4(b)]	against low [3(c)]
$R(w) = 1$	against high [3(d)]	unbiased [3(b)]	against low [3(c)]
$R(w) > 1$	against high [3(d)]	against high [4(a)]	against low [3(c)]

Table 1: Equilibrium bias for different penalties and risk preferences

economic means. As noted in the introduction, the result that jail overdeters the rich and underdeters the poor is common in the theoretical literature. Empirically, research by Grasmick and Bryjak (1980) shows that individuals generally differ in how severe they perceive various forms of criminal sanctions. Unfortunately, their study does not relate perceived severity to economic covariates. Other papers have demonstrated a negative relationship between wages and employment on the one hand, and criminal participation on the other (e.g. Myers, 1983; Grogger, 1998; Gould et al., 2002). However, these studies leave unanswered the question whether this relationship is due to deterrence effects.[10] Bar-Ilan and Sacerdote (2004) examine the response of red-light violations to fine increases in Israel and San Francisco. The study finds that the elasticity of violations to fines is larger for younger drivers and owners of older cars, both of which are correlated with the driver's affluence.[11]

4 Social Norms and Statistical Discrimination

The previous section offered an explanation for why poorer individuals may be treated less favorably than richer ones in the courts. If income or wealth is correlated with

---

[10]There are many competing explanations. Low-wage earners may offend more because they have more to gain from crime, especially property crime. Similarly, unemployed individuals may participate in crime at a higher rate because they have more time to do so.

[11]As far as prison is concerned, anecdotal evidence suggests that its deterrence effect is large for the very affluent, and small or even negative for the very poor. For example, the 2002 Sabanes-Oxley Act increased the potential prison sentences for certain types of accounting fraud, a crime typically committed by affluent offenders undeterred by financial penalties. At the other end of the spectrum, on May 29, 2008, West Palm Beach, Florida, NBC affiliate WPTV reported on a homeless man who committed a drug crime in order to be sent to jail. The report quoted the man saying "I feel so much more free here. You can't be free out there." (See `http://www.wptv.com/news/local/story.aspx?content_id=3f6063d4-b79d-4f8f-8d78-2103dc8998ba`.)

other characteristics, such as a person's race, then the story we told can explain why racial groups with poorer members may also be the ones whose members are more easily convicted of crimes.[12] In this section, we dispose of the assumption that individuals differ in terms of their income ex-ante. We show that prejudice against certain identifiable subgroups of the population may nevertheless exist.

4.1   An extended model

In our extended model, there are $L \geq 2$ subgroups of the population. Membership in one group is a label $l \in \{1, \dots, L\}$ that is publicly observable but a priori without economic significance. Thus, in contrast to the previous section, all agents are ex-ante alike with respect to such characteristics as their productivity, human capital, etc., which may influence a person's income. We assume instead that differences in the realized value of an agent's income are the direct result of the bias held against the agent's group in the justice system. In particular, we let $v : [0, 1] \to [\underline{w}, \overline{w}]$ ($\underline{w} < \overline{w} < \infty$) describe an agent's income as a function of the prejudice held against the agent.

We assume that v is continuous and weakly decreasing. This reflects the idea that it may be harder for a person to find a job, or a high-paying job, if this person is assumed to be more prone to criminal behavior. This can be for several reasons. First, criminal activity may adversely affect an individual's productivity in a job, for example because of increased absenteeism. Second, criminal activity may be targeted at the employer directly, for instance by stealing from a job site. Third, an employer may liable for an employee's criminal actions when the employee harms a customer or another employee (under the legal theories of respondeat superior and negligent hiring).

There is empiricial evidence that a feedback channel from (perceived) criminality to labor market outcomes indeed exists. Numerous studies find a negative impact of observed prior convictions on labor market outcomes, suggesting that employers are adverse to hiring convicted criminals (e.g. Lott, 1992; Grogger, 1995; Nagin and Waldfogel, 1998; and many others). Other papers measure the effects of perceived criminality toward minorities on labor market outcomes. Bushway (2004) demonstrates that the black-white wage ratio is higher in states where criminal records databases are more easily accesible to employers. Holzer et al. (2006) show that employers who routinely conduct criminal background checks of applicants are more likely to hire African-American applicants than employers who do not perform routine checks (after controlling for characteristics of each firm's applicant pool). Both studies suggest that in the absence of information about an individual's criminal history group characteristics such as race may be used by employers to make inferences about an applicant's criminality, and that such inferences are indeed used in hiring or pay decisions.

---

[12]Even if the income distribution is the same within all racial groups, it is possible that differences in criminal participation and convictions emerge: If there are multiple prejudice levels consistent with a particular income level (which is possible if the probability of accidental investigation is high enough), then a person's race may serve as a coordination device by which a particular equilibrium is selected.

4.2   Equilibrium

Since the only observable difference between individuals is that they belong to different subgroups of the population, we now replace the prejudice $p(w)$ against persons of income w by a prejudice p_l against persons who belong to group l. Similarly, replace $q(w)$ by q_l, and $t(w)$ by t_l. An equilibrium in this case is now a collection $(w_l^*, p_l^*, q_l^*, t_l^*)_{l \in \{1,\dots,L\}}$ such that for all $l \in \{1, \dots, L\}$,

$$p_l^* = 1 - Q(q_l^*), \tag{10}$$

$$q_l^* = [1 - F(t_l^*) - \lambda(1 - G(t_l^*))]\, \Delta_l, \tag{11}$$

$$\theta(p_l, t_l^*) = \alpha, \tag{12}$$

$$w_l^* = v(p_l^*), \tag{13}$$

where $\Delta_l = u(w_l^*) - u(w_l^* - \rho(w_l^*))$ is defined as before. This is essentially the same set of defining equations as (2)–(4), except that a fourth condition has been added (condition (13)). This condition states that the equilibrium income level of group l, w_l^*, must be consistent with the prejudice level p_l^* held against group l. We now say an equilibrium is *biased* if there exist two different groups l and l' such that $p_l^* > p_{l'}^*$. If such real inter-group differences arise in equilibrium, then these outcomes are "sunspot equilibria." An equilibrium is *unbiased* if $p_1^* = \dots = p_L^*$.

Equilibria can be derived from the intersection of $p^*(w)$ defined through (2) with the income function $v(p)$. Let

$$\mathcal{S} = \{(p, w) \in [0, 1] \times [\underline{w}, \overline{w}] : p = p^*(v(p))\}$$

be the set of all intersecting points of the two curves. An equilibrium in the extended model can then be constructed by assigning to each group $l \in \{1, \dots, L\}$ an element in $\mathcal{S}$, corresponding to the prejudice p_l^* against group l and the income level w_l^* of its members. If $\mathcal{S} \neq \emptyset$ an equilibrium exists, in which case there must be an unbiased equilibrium, as all groups can be assigned the same (p, w)-pair. However, if $\mathcal{S}$ contains more than one element there are also biased equilibria, as any assignment of groups to (p, w)-pairs contained in $\mathcal{S}$ represents an equilibrium in the extended model. We have the following result

Lemma 5. *Regardless of the punishment used, if $0 < \lambda \leq \overline{\lambda}$ there exists an equilibrium in the extended model. Furthermore,*

   (a) if Δ strictly decreases in w or is constant, then the equilibrium is unique and unbiased, and

   (b) if Δ increases in w, biased equilibria exist for certain income functions v.

   We illustrate Lemma 5 in Figure 1. Note that the income function v that is depicted in both diagrams exhibits a drop at an interior p, and this will play a role for the

multiplicity of equilibria in case (b) of the result. Such a drop can happen if there are two main job types that pay different wages—managerial and simple jobs, say—and each employer has a threshold for p at which they are willing to hire workers into the high-wage job. If the job-specific wages and prejudice thresholds are similar across firms, then the resulting income function will look like the one in Figure 1.
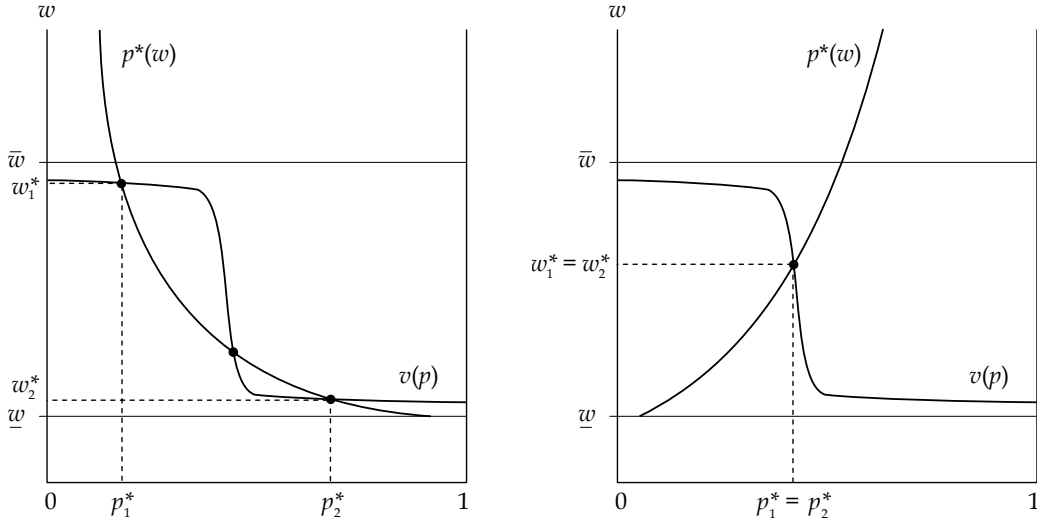

Figure 1: Biased and unbiased equilibrium in the extended model

The left diagram depicts the case of Lemma 5 (b), which applies to prison sentences. A stereotyping equilibrium can arise, in which different identifiable subgroups face an unfavorable bias in the courts. In this case, there are two groups: Group 1 earns a relatively high income w_1^* and faces a relatively low prejudice p_1^*. The opposite holds for group 2. In contrast, the right diagram shows the same v-curve as before, but a simple fine is used instead of prison to punish convicted agents. This corresponds to Theorem 2 (a). It is easy to see that there cannot be multiple intersections of p^* and v now, so that the only equilibrium is an unbiased one.

4.3   Example: A dynamic link from prejudice to income

In this section, we provide an example that demonstrates how a feedback channel from prejudice to income may arise within the criminal justice system. That is, we do not rely on labor market interactions to generate discriminatory outcomes. By using expected lifetime earnings as the relevant income variable, we show that the sanctions imposed by the courts can be directly responsible for differences in lifetime earnings.

Consider the following stylized environment. Agents live for infinitely many periods. At the beginning of each period, an agent works and earns a fixed wage of y, which is the same for all individuals, unless the individual is imprisoned and earns zero. All

income is consumed in the period it is earned, and at the end of each period the events described in Section 2 unfold: Agents observe their i.i.d. η-shocks, decide whether to commit the crime, and possibly stand trial. If an agent is convicted (rightfully or wrongfully), he is removed from employment and sentenced to life in prison, representing a permanent reduction in income. Otherwise, the agent starts the next period as a working individual. In computing their expected lifetime earnings, individuals apply a discount factor $\beta < 1$ and do not include future realizations of η.[13] At the time the agent decides whether to commit a crime, the expected lifetime consumption for an agent from the *next* period on, conditional on entering the next period as a free individual, is

$$v(p) = \frac{\beta}{1 - \beta\xi(q,t)}y,$$

where $\xi(q,t) = [1 - Q(q)]\, F(t) + Q(q)\,[1 - \lambda(1 - G(t))]$ is the period-to-period "survival probability" associated with the tuple (q,t). The deterrence effect of the punishment is therefore the prospect of earning zero from the next period onward if convicted, instead of earning an expected continuation utility $v(p)$. We hence set $\Delta = v(p) - 0 = y \cdot \beta/(1 - \beta\xi(q,t))$ and compute $v(p)$ and $p^*(w)$ for the following parameter values:[14]

$$y = 1,\ \alpha = 0.95,\ \beta = 0.95,\ \lambda = 0.01,\ \eta \sim U[0,5],\ F(t) = t^2,\ G(t) = 2t - t^2.$$

The result is plotted in Figure 2. One can see that there are in fact three intersections of the two curves. Thus, if there are two or more racial (or otherwise distinguishable) groups in the population, whose members all earn $y = 1$ when not incarcerated, biased equilibria can arise simply because two different groups can be "assigned" different prejudice-lifetime income pairs which correspond to the intersections in Figure 2.

This dynamic model can explain why some individuals choose to live a "life of crime" and why this choice may be correlated with characteristics such as race. Consider an individual in the high crime/low lifetime income group. Each time he decides whether to commit a crime he compares the benefit (η), which is distributed equally across the entire population, against the expected cost. The expected cost is that the individual (with some probability) loses his criminal career and goes to jail. However, continuing a life of crime is not a very enticing prospect either, because a career criminal expects to be jailed sooner or later anyways. Such a person is therefore less likely to be deterred by this prospect. The opposite holds for the choice to live a low-crime life, and the usual stereotyping argument can be made to sort individuals into different such equilibria based on some observable characteristic. The crucial aspect here is that, because

---

[13]Unless η is literally regarded as the material benefit from a crime (for instance money stolen), this assumption seems not unreasonable. For offenses such as drug consumption, it seems indeed natural to impose a strong bias for the presence regarding the benefit η.

[14]This is not entirely in line with some of our previous assumptions; for instance $f(0) = 0$ and η is bounded in this example. However, these assumptions are sufficient to ensure equilibria existed but not necessary, and our example illustrates that the same biased outcomes can also arise in other cases.
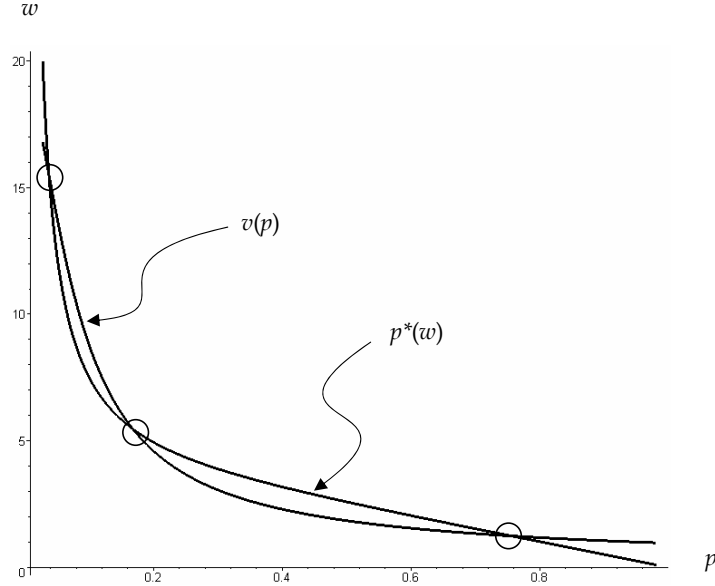
Figure 2: Numerical example, where $v(p)$ is expected discounted lifetime income

group membership cannot be altered, it is not possible for individuals to break out of the high-crime equilibrium—they would be treated in an adverse manner by the courts even if they decided not to commit crime ever.

Finally, note that in this example observed wages are entirely uncorrelated with crime and conviction rates, while judicial stereotyping based on race still persists. This is interesting, as several empirical studies find that in the U.S. race is predictive of the incarceration rate even after controlling for economic factors (e.g. Bjerk, 2006; Krivo and Peterson, 1996; Raphael and Winter-Ebmer, 2001; and Trumbull, 1989).

5 Discussion and Policy Implications

This paper demonstrates how statistical discrimination can result in the judicial system in much the same way as economists have considered it in the labor market. Our results yield a number of policy implications and suggest several extensions worthwhile pursuing. Below, we discuss some of them.

Labor market interactions. It is worth discussing the linkage between discrimination in the criminal justice system and the labor market. As shown in Section 3, lower incomes for certain groups can lead to increased incentives to commit crime. Such labor market differences can themselves be the result of (past) discrimination in the labor market. People who observe the increased crime rate among such groups may conclude that members of these groups are inherently more likely to commit crime, rather than attribute the increased crime to lower incomes. It is conceivable that such

a prejudice can become self-fulfilling, even under policies which aim to eliminate these income differences across groups (see, for example, Sah, 1991). Thus, discrimination in the legal system which originally arose from discrimination in the labor market would be unaffected by policies such as affirmative action.

Monetary vs. non-monetary sanctions. In order to eliminate prejudice in the legal system, policies targeted at the courts can be employed. Obviously, the equivalent policy to affirmative action—setting conviction quotas—would not be appealing. However, we show that prejudices can be eliminated through the choice of punishment. Biased equilibria arise when penalties are solely in terms of jail time, and the amount of time is independent of the defendant's type. On the other hand, fines and in particular means-adjusted fines can avoid this problem. Of course, in order to achieve a desired deterrence effect through fines, it is possible that the fine must be chosen so high that some defendants would not be able to pay it (defendants are judgement proof). In such a case, penalties that first confiscate an individual's wealth and then assign jail time based on the amount of fine left to pay would eliminate such equilibria. This form of penalty is prescribed by Polinsky and Shavell (1984, 2000). Thus, the question as to why jail (without first confiscating wealth) is used so frequently as a punishment remains.

The role of information at trials. The discriminatory outcomes we explore in Section 4 are the result of a "sunspot effect," where labels such as race are used as coordination devices. It is important to notice that the criminal justice system itself can generate such sunspots through its record-keeping technology. A case in point is whether or not juries should be informed of a defendant's prior convictions. If this characteristic is observable to the jury, it can act as a coordination device. That is, there can be a social expectation that persons convicted before are more likely to commit further crimes. This expectation can become self-fulfilling, for reasons similar to the ones explored in Section 4.3. Moreover, this belief may be grounded in an entirely different theory, namely that some individuals are inherently more prone to commit crime than others.[15] The obvious way to eliminate this bias is to restrict information about prior convictions at trials. If this is not possible, then a schedule of graduated penalties for repeat offenders (e.g. Emons, 2007) may be desirable, as it increases the deterrence effect for individuals subject to an unfavorable prejudice (i.e., individuals with prior convictions).

The effects of policing. In this paper we do not formally consider the process by which suspects are apprehended and evidence is generated. Such an analysis could be an interesting extension of the model we present. That is, one could investigate how the results change when the evidence generating functions F and G are altered.

---

[15]While this may or may not be the case, our argument is that the statistical discrimination theory is observationally equivalent.

For example, if an increase in policing were to increase (proportionally) the evidence generated for both innocent and guilty parties, then an increase in expenditures on police would have no effect on the existence of a biased equilibrium. If, however, an increase in policing increased evidence against guilty parties and decreased (or had no effect) on evidence against innocent parties, then an increase in policing could eliminate stereotyping equilibria. In order to explore the "true" properties of evidence generation, empirical studies of the effect of policing (e.g. Knowles, Persico and Todd, 2001) could perhaps be used. However, since it cannot be known with certainty if convicted parties are actually innocent or guilty, it is unclear how much light such studies can ultimately shed on this issue.

Appendix

We start with two preliminary observations. First, note that $1 - F(t) - \lambda(1 - G(t))$ is weakly decreasing in t if and only if $\lambda \leq f(t)/g(t)$. Since $f(t)/g(t)$ is increasing by assumption, if $\lambda \leq \bar{\lambda}$ then $1 - F(t) - \lambda(1 - G(t))$ is non-increasing. For $\lambda \leq \bar{\lambda}$ we therefore get

$$\lambda \leq \bar{\lambda} \quad \Rightarrow \quad \frac{\partial}{\partial t} q = [-f(t) + \lambda g(t)] \Delta \leq 0. \tag{14}$$

Thus, if $\lambda \leq \bar{\lambda}$, an increase in the conviction threshold t indeed leads to a decrease in the decision threshold q of the agent. Second, since $1 - F(1) - \lambda(1 - G(1)) = 0$, we have

$$\lambda \leq \bar{\lambda} \quad \Rightarrow \quad \frac{\partial}{\partial \Delta} q = 1 - F(t) - \lambda(1 - G(t)) = - \int_t^1 [-f(s) + \lambda g(s)] \, ds \geq 0. \tag{15}$$

Thus, if $\lambda \leq \bar{\lambda}$, an increase in the potential penalty Δ leads to an increase in the decision threshold q of the agent.

Proof of Lemma 1. To prove an equilibrium exists, we make a standard fixed point argument. Fix any w and define three maps,

$$\mathcal{T}_1 \quad : \quad q \to p \; : \; [0, \infty) \to [0, 1],$$

$$\mathcal{T}_2^w \quad : \quad t \to q \; : \; [0, 1] \to [0, \infty),$$

$$\mathcal{T}_3 \quad : \quad p \to t \; : \; (0, 1) \to [0, 1]$$

by (2), (3) (given w), and (4), respectively; these are all given in Section 2.4. Note that $\mathcal{T}_1$, $\mathcal{T}_2^w$ and $\mathcal{T}_3$ are continuous on their respective domains. $\mathcal{T}_3$ is well-defined through (4) on $(0, 1)$ only; however it can be extended continuously to $[0, 1]$ by setting $\mathcal{T}_3(0) = 1$ and $\mathcal{T}_3(1) = 0$. Further, as $\mathcal{T}_2^w$ is continuous on a compact domain, its image is bounded. We can hence restrict the range of $\mathcal{T}_2^w$, as well as the domain of $\mathcal{T}_1$, to $[0, \hat{q}(w)]$ for sufficiently large $\hat{q}(w)$. Now define a new map

$$\mathcal{T}^w \; : \; [0, 1] \times [0, \hat{q}(w)] \times [0, 1] \to [0, 1] \times [0, \hat{q}(w)] \times [0, 1]$$

by setting $\mathcal{T}^w(p,q,t) = (\mathcal{T}_1(q), \mathcal{T}_2^w(t), \mathcal{T}_3(p))$. Since $\mathcal{T}^w$ is continuous and maps a compact subset of $\mathbb{R}^3$ into itself, we can apply Brouwer's fixed point theorem to show, for given w, there exists $(p^*(w), q^*(w), t^*(w))$ such that

$$(p^*(w), q^*(w), t^*(w)) = \mathcal{T}^w(p^*(w), q^*(w), t^*(w));$$

thus it solves (2)–(4) simultaneously. Since such a fixed point can be constructed for each w independently, an equilibrium as defined above exists.

To prove uniqueness, let $\lambda \le \bar{\lambda}$ and suppose there are two equilibria, $(p^*, q^*, t^*) \ne (\widetilde{p}^*, \widetilde{q}^*, \widetilde{t}^*)$. Thus there exists w such that $q^*(w) \ne \widetilde{q}^*(w)$ (otherwise $p^*(w) = \widetilde{p}^*(w)$ $\forall w$ by (2), which implies $t^*(w) = \widetilde{t}^*(w)$ $\forall w$, by (4), but then the equilibrium would be unique). So suppose, without loss of generality, that $q^*(w) > \widetilde{q}^*(w)$ for some w. Condition (2) then implies $p^*(w) < \widetilde{p}^*(w)$, and using (4) we have $t^*(w) > \widetilde{t}^*(w)$. If $\lambda \le \bar{\lambda}$ then using (14) we get $q^*(w) \le \widetilde{q}^*(w)$, a contradiction. Hence the equilibrium is unique if $\lambda \le \bar{\lambda}$. $\qquad\square$

Proof of Lemma 2. Let $w > w'$ and suppose $\Delta(w) > \Delta(w')$. If $p(w) \ge p(w')$, then by condition (2), $q(w) \le q(w')$. However, by condition (4) and the fact that f/g increases, we have $t(w) \le t(w')$. Thus if $\Delta(w) > \Delta(w')$ then (14)–(15) imply $q(w) > q(w')$, which is a contradiction, and therefore $p(w) < p(w')$. From (4) it follows then that $t(w) > t(w')$, and from (2) it follows that $q(w) > q(w')$. Exactly the opposite argument can be made when $w > w'$ and $\Delta(w) < \Delta(w')$. Finally, when Δ is a constant then (3) is independent of w so that p^*, q^*, and t^* are constant and hence constitute an unbiased equilibrium. $\qquad\square$

Proof of Theorem 3. If $u''(w) = 0$, then $\Delta'(w) = u'(w) - u'(w - \delta) = 0$, and Lemma 2 (iii) implies (a). If $\rho(w) = \gamma w$ and $u(w) = a \ln w + b$, then $\Delta'(w) = a/w - a/w = 0$, and Lemma 2 (iii) implies (b) as well. In case of imprisonment, $\Delta(w) = u(w) - u(w_0)$, which is strictly increasing in w since $u'(w) > 0$, and applying Lemma 2 (ii) yields (c). In case of a simple fine, $\Delta(w) = u(w) - u(w - \delta)$, which is strictly decreasing in w if $u''(w) < 0$ $\forall w$. Thus, applying Lemma 2 (i) yields (d). $\qquad\square$

Proof of Theorem 4. Suppose $R(w) > 1$ for all w, or equivalently $-\gamma w u''(\gamma w)/u'(\gamma w) > 1$ $\forall \gamma$, and thus

$$\frac{\partial}{\partial \gamma} \left[\gamma u'(\gamma w) \right] = u'(\gamma w) + \gamma w u''(\gamma w) < 0 \ \ \forall \gamma. \tag{16}$$

Let $\mu(w) = w - \rho(w)$ be the income left to the individual after the fine $\rho(w)$. Since $\mu(w) < w$, (16) implies

$$\frac{\mu(w)}{w} u'(\mu(w)) \ = \ u'(w) - \int_{\mu(w)/w}^1 \gamma u'(\gamma w) \, d\gamma \ > \ u'(w). \tag{17}$$

Suppose now that $\varepsilon(w) \le 1$ $\forall w$: $w\rho'(w)/\rho(w) \le 1$. Multiplying this inequality by $\rho(w)/w$ yields $\rho'(w) \le \rho(w)/w$, and expressing the fine as $\rho(w) = w - \mu(w)$ we get

18

$\mu'(w) \geq \mu(w)/w$. Then by (17)

$$\Delta'(w) \;=\; \frac{\partial}{\partial w}\left[u(w) - u(\mu(w))\right] \;=\; u'(w) - \mu'(w)u'(\mu(w))$$

$$\leq\; u'(w) - \frac{\rho(w)}{w}u'(\mu(w)) \;<\; 0.$$

By Lemma 2 (i), therefore, the equilibrium is biased against higher types. Analogous steps can be repeated for $R(w) < 1$ and $\varepsilon(w) \geq 1$, in which case $\Delta'(w) > 0$ and the equilibrium is biased against lower types by Lemma 2 (ii). $\qquad\square$

Proof of Lemma 5. We first prove the existence part. Note that v is a continuous function mapping $p \in [0,1]$ to $w \in [\underline{w}, \overline{w}]$. If $\lambda \leq \overline{\lambda}$, $p^*(w)$ is a function by Lemma 1. As Δ is continuous in w regardless of the punishment used, w enters the mapping $\mathcal{T}$ defined in the proof of Lemma 1 continuously, so p^* is continuous. This implies that in $[0,1] \times [\underline{,}\overline{w}]$, the graphs of v and p^* intersect at least once, so $\mathcal{S} \neq \emptyset$.

We now prove statements (a) and (b) of the result. For (a) note that by Lemma 2 (i) p^* strictly increases for strictly decreasing Δ. Therefore, for each v there is exactly one $w \in [0,1]$ such that $v(p^*(w)) = w$. If Δ is constant, then by Lemma 2 (iii) p^* is constant. It will hence become a vertical line in p-w space, which is intersected by any decreasing, continuous v exactly once. Hence $|\mathcal{S}| = 1$, so that exactly one equilibrium exists, which must be unbiased. For (b), note that by Lemma 2 (ii) p^* strictly decreases for strictly increasing Δ. Therefore, there exists a continuous, decreasing function v for which the following holds: There exists $w_1, w_2 \in [\underline{w}, \overline{w}]$, $w_1 \neq w_2$, such that $v(p^*(w_1)) = w_1$ and $v(p^*(w_2)) = w_2$. For such v, $|\mathcal{S}| > 1$, and a biased equilibrium exists. $\qquad\square$

References

ALEXEEV, M. AND J. LEITZEL (2004): "Racial Profiling," mimeo, Indiana University.

ARROW, K. (1973): "The Theory of Discrimination," in: O. Ashenfelter and A. Rees (eds.): *Discrimination in Labor Markets*, Princeton University Press, 3–33.

BAR-ILAN, A. AND B. SACERDOTE (2004): "The Response of Criminals and Noncriminals to Fines," *Journal of Law and Economics*, **47**, 1–17.

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

BJERK, D. (2006): "Theory and Evidence Regarding the Effects of Segregation on Crime Rates," mimeo, Claremont McKenna College.

BJERK, D. (2007): "Racial Profiling, Statistical Discrimination, and the Effect of a Colorblind Policy on the Crime Rate," *Journal of Public Economic Theory*, **9**, 425–545.

BURKE, A. (2007): "Neutralizing Cognitive Bias: An Invitation to Prosecutors," *NYU Journal of Law and Liberty*, **2**, 512–530.

BUSHWAY, S. (2004): "Labor Market Effects of Permitting Employer Access to Criminal History Records," *Journal of Contemporary Criminal Justice*, **20**, 276–291.

COATE, S. AND G. LOURY (1993): "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, **83**, 1220–1240.

EMONS, W. (2007): "Escalating Penalties for Repeat Offenders," *International Review of Law and Economics*, **27**, 170–178.

GEORGAKOPOULOS, N. (2004): "Self-fulfilling Impressions of Criminality: Unintentional Race Profiling," *International Review of Law and Economics*, **24**, 169–190.

GOULD, E., B. WEINBERG, AND D. MUSTARD (2002): "Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997," *The Review of Economics and Statistics*, **84**, 45–61.

GRASMICK, H. AND G. BRYJAK (1980): "The Deterrent Effect of Perceived Severity of Punishment," *Social Forces*, **59**, 471–491.

GROGGER, J. (1995): "The Effect of Arrests on the Employment and Earnings of Young Men," *Quarterly Journal of Economics*, **110**, 51-71.

GROGGER, J. (1998): "Market Wages and Youth Crime," *Journal of Labor Economics*, **16**, 756-791.

HILLSMAN, S. (1990): "Fines and Day Fines," *Crime and Justice*, **12**, 49–98.

HOLZER, H., S. RAPHAEL, AND M. STOLL (2006): "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers," *Journal of Law and Economics*, **49**, 451–480.

KNOWLES, J., N. PERSICO, AND P. TODD (2001): "Racial Bias in Motor-Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, **109**, 203–229.

KRIVO, L. AND R. PETERSON (1996): "Extremely Disadvantaged Neighborhoods and Urban Crime," *Social Forces*, **75**, 619–650.

LEVITT, S. (1997): "Incentive Compatibility Constraints as an Explanation for the Use of Prison Sentences Instead of Fines," *International Review of Law and Economics*, **17**, 179–192.

LOTT, J. (1987): "Should the Wealthy Be Able to 'Buy Justice'?" *Journal of Political Economy*, **95**, 1307–1316.

LOTT, J. (1992): "An Attempt at Measuring the Total Monetary Penalty from Drug Convictions: The Importance of an Individual's Reputation," *Journal of Legal Studies*,

11, 159–187.

MORRIS, N. AND M. TONRY (1990): *Between Prison and Probation*, Oxford University Press, Oxford.

MYERS, S. (1983): "Estimating the Economic Model of Crime: Employment Versus Punishment Effects," *Quarterly Journal of Economics*, **98**, 157–166.

NAGIN, D. AND J. WALDFOGEL (1998): "The Effect of Conviction in Income Through the Life Cycle," *International Review of Law and Economics*, **18**, 25–40.

PERSICO, N. (2002): "Racial Profiling, Fairness, and Effectiveness of Policing," *American Economic Review*, **92**, 1472–1497.

PHELPS, E. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, **62**, 659–661.

POLINSKY A. AND S. SHAVELL (1984): "The Optimal Use of Fines and Imprisonment," *Journal of Public Economics*, **24**, 89–99.

POLINSKY A. AND S. SHAVELL (2000): "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature*, **38**, 45–76.

POSNER, R. (1992): *Economic Analysis of Law*, 4th Edition, Little, Brown and Company, Boston.

RAPHAEL, S. AND R. WINTER-EBMER (2001): "Identifying the Effect of Unemployment on Crime," *Journal of Law and Economics*, **44**, 259–283.

SAH, R. (1991): "Social Osmosis and Patterns of Crime," *Journal of Political Economy*, **99**, 1272–1295.

SPIEGEL, U. AND J. TEMPLEMAN (1989): "Economics of Discriminatory Sentencing," *Journal of Quantitative Criminology*, **5**, 317–332.

TRUMBULL, W. (1989): "Estimations of the Economic Model of Crime Using Aggregate and Individual Level Data," *Southern Economic Journal*, **56**, 423–439.

VERDIER, T. AND Y. ZENOU (2004): "Racial Beliefs, Location, and the Cause of Crime," *International Economic Review*, **45**, 731–760.