

3-2003

# Can Racially Unbiased Police Perpetuate Long-Run Discrimination?

Helle Bunzel

*Iowa State University*, [hbunzel@iastate.edu](mailto:hbunzel@iastate.edu)

Philippe Marcoul

*Iowa State University*

Follow this and additional works at: [http://lib.dr.iastate.edu/econ\\_las\\_workingpapers](http://lib.dr.iastate.edu/econ_las_workingpapers)



Part of the [Other Economics Commons](#), [Race and Ethnicity Commons](#), and the [Social Control, Law, Crime, and Deviance Commons](#)

---

## Recommended Citation

Bunzel, Helle and Marcoul, Philippe, "Can Racially Unbiased Police Perpetuate Long-Run Discrimination?" (2003). *Economics Working Papers (2002–2016)*. 212.

[http://lib.dr.iastate.edu/econ\\_las\\_workingpapers/212](http://lib.dr.iastate.edu/econ_las_workingpapers/212)

This Working Paper is brought to you for free and open access by the Economics at Iowa State University Digital Repository. It has been accepted for inclusion in Economics Working Papers (2002–2016) by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Can Racially Unbiased Police Perpetuate Long-Run Discrimination?

## **Abstract**

We develop a stylized dynamic model of highway policing in which a non-racist police officer exhibits a cognitive bias: relative overconfidence. The officer is given incentives to arrest criminals but faces a per stop cost which increases when the racial mix of her stops differs from that of the population. Every period, she observes the racial composition of jail inmates (generated from arrests made by her peers) and forms estimates about the crime rates of each race. In some settings, her overconfidence leads her to overestimate the crime rate of one race relative to another causing the long-run racial composition of the jail population to deviate from the “fair” one (one where the racial mix in jails is identical to that in the criminal population). We compare this to a situation where officers have detailed stop data on each race, similar to data being currently collected in many US states.

## **Disciplines**

Other Economics | Race and Ethnicity | Social Control, Law, Crime, and Deviance

# IOWA STATE UNIVERSITY

**Can Racially Unbiased Police Perpetuate Long-Run  
Discrimination?**

Helle Bunzel, Philippe Marcoul

March 2003

Working Paper # 03004

**Department of Economics  
Working Papers Series**

**Ames, Iowa 50011**

Iowa State University does not discriminate on the basis of race, color, age, religion, national origin, sexual orientation, gender identity, sex, marital status, disability, or status as a U.S. veteran. Inquiries can be directed to the Director of Equal Opportunity and Diversity, 3680 Beardshear Hall, (515) 294-7612.

# Can Racially Unbiased Police Perpetuate Long-Run Discrimination?

HELLE BUNZEL\*                      PHILIPPE MARCOUL  
IOWA STATE UNIVERSITY            IOWA STATE UNIVERSITY

January 2006<sup>†</sup>

ABSTRACT. We develop a stylized dynamic model of highway policing in which a non-racist police officer exhibits a cognitive bias: relative overconfidence. The officer is given incentives to arrest criminals but faces a per stop cost which increases when the racial mix of her stops differs from that of the population. Every period, she observes the racial composition of jail inmates (generated from arrests made by her peers) and forms estimates about the crime rates of each race. In some settings, her overconfidence leads her to overestimate the crime rate of one race relative to another causing the long-run racial composition of the jail population to deviate from the “fair” one (one where the racial mix in jails is identical to that in the criminal population). We compare this to a situation where officers have detailed stop data on each race, similar to data being currently collected in many US states.

## 1. INTRODUCTION

The words “discrimination” and “prejudice” are fairly tightly interlinked; in fact a standard dictionary definition of the former is “unfair treatment of a person or group on the basis of prejudice”. The “neoclassical models of discrimination” due to Becker (1971) argue that a major source of discrimination is personal prejudice: People act as if they have a taste for discrimination. A powerful result that emerges from this paradigm is the following: Since discrimination is potentially costly, if competitive forces are at work, the discriminators would be punished and not survive in the long run. Subsequent to Becker (1971), much work has focused on statistical prejudice or “statistical discrimination” which does not arise from prejudice, but rather when certain perceived group characteristics are statistically projected onto individuals. Standard labor economics textbooks blame statistical

---

\*Corresponding author. Department of Economics, 260 Heady Hall, Iowa State University, Ames IA 50011. Phone (515) 294 6163, Fax: (515) 294 6644, e-mail: hbunzel@iastate.edu

<sup>†</sup>We are grateful to Joydeep Bhattacharya, Subir Bose, Jacques Crémer, Preston McAfee, and Brent Kreider for extensive comments on earlier versions of this paper, Alicia Carrquiry, Jeremy Greenwood, Saul Kassin, Nicola Persico, Rajesh Singh and Oscar Volij for discussions, and participants at the 8th Spring Meeting of Young Economists in Leuven, Belgium as well as the NFL workshop at Iowa State University in general and David Frankel in particular, for extensive feedback. As usual, all remaining errors are our own.

discrimination on the tests used to screen people: maybe the test underpredicts for one group, or even when it predicts correctly on average, it does a much poorer job of predicting for one group as opposed to another. Unlike prejudicial discrimination, statistical discrimination may survive even in the long run (Moro and Norman, 2004).

This paper explores one possible way in which statistical discrimination may arise and persist in an environment where the tests are not at fault. To that end, it borrows from a burgeoning literature in economics and psychology on heuristics and cognitive biases and shows how statistical discrimination may be influenced by these. According to psychologists, cognitive biases appear when people make probabilistic judgements but their inferences violate the very core of rationality: Bayes rule. We focus our discussion using a bias that was discussed even by Adam Smith: overconfidence. Formally, relative overconfidence bias is the practice of systematically overestimating the accuracy of one's decisions and the precision of one's knowledge relative to peers.<sup>1</sup> This paper studies the role of a relative overconfidence bias in the specific context of racial profiling and persistent racial discrimination. It lays out the structure of a model in which police officers have no innate prejudice against a race and yet "unfair" racial discrimination emerges because police officers suffer from a relative overconfidence bias while receiving incentives to catch criminals.<sup>2</sup>

Psychologists have a long tradition of studying and documenting that humans in general suffer from overconfidence (for a survey see Yates (1990)). They also find that people were more confident of their predictions in fields where they have self-declared expertise (Heath and Tversky 1991). There is an extensive literature in psychology documenting the fact that police officers are in fact overconfident. Specifically numerous studies document that police officers perform no better than laymen at detecting criminals during interviews, but that their confidence in their judgement is inflated, see for example DePaulo and Pfeifer (1986), Meissner and Kassin (2002) and Garrido, Masip and Herrero (2004). In addition, Mullin (1995) reviews a study of Scottish police which finds "detective Sergeants, as a group, think they know a lot more than they actually do!"

---

<sup>1</sup>This phenomenon is well-documented and is believed to be widespread in society. For example, Babcock and Loewenstein (1997) write, "well over half of survey respondents rate themselves in the top 50 percent of drivers, ethics, managerial prowess, productivity, and a variety of desirable skills." In a more recent study, Glazer and Weber (2005) offer empirical evidence that relative overconfidence matters; they find that investors who believe they are above average in terms of investment skills trade more frequently.

<sup>2</sup>Note that there are many other types of cognitive biases we could have chosen to examine in this context: Imperfect recall, moral overconfidence (as suggested by a referee) and confirmatory bias to mention a few. Relative overconfidence seems to be a natural place to start because it is well-documented, even among police officers.

The use of racial profiling by the police as a crime-prevention tool has been a topic of debate for several years now. It first became a national issue in the media when it was suspected that highway troopers stopping cars on the I-80 corridor were making heavy use of race as a proxy for possible criminality. By now, it has become generally accepted that at least some racial profiling is practiced by police around the country. Those who oppose racial profiling as a crime-fighting tool find it unreasonable to simply use race, without any additional indication of criminal behavior, as probable cause. Others argue that if members of one race are statistically more likely to be involved in a certain crime, racial profiling has to be an important tool when fighting crime.

We examine the issues surrounding racial profiling as a crime-prevention tool in the context of “high discretion interdiction” (Persico, 2002). We will interpret our model in the context highway stops.<sup>3</sup> We assume that police officers are not racist; they only suffer from a relative overconfidence bias. The relative overconfidence takes the following form: Any officer believes she is better at apprehending criminals than all other officers. Specifically, while in actuality any officer apprehends all the criminals she encounters, she assumes all *other* officers only manage to incarcerate a fraction of the criminals they interdict, presumably because they bungle up crime investigations etc.

We assume, without loss of generality, that the percentage of criminals is marginally higher among blacks. In addition, we assume that the police do not know these percentages, and use the racial distribution of jail inmates to estimate them. The officers are given incentives to arrest as many guilty people as possible, but they also face a per stop cost which increases whenever the proportion of people they stop from one racial group deviates from the proportion of this group in the whole population. The cost is imposed equally regardless of which race is oversampled. This seems reasonable in a world where there is outrage over every alleged instance of reported racial profiling and where the public’s faith in the police is low, especially among minorities. We posit that individual officers, when deciding whom to stop, care only about maximizing their revenue. In this framework, we show that there is long-run excess incarceration of the more criminal race, compared to their true share of the criminal population. In other words, if 11% of whites commit crimes, but only 10% of blacks do, in the long run, whites will comprise *more* than 11/21 of the incarcerated population. In that sense, jail rates will be “unfair”.

The assumed relative overconfidence bias of the police affects this result in the following manner. An officer maximizes her revenue by equating the marginal benefit of a stop to

---

<sup>3</sup>Traffic stops make up 52% of citizen interaction with police in the US (Langan et al, 2001).

the marginal cost of deviating from strict random sampling. The bias makes her think that a fraction of the criminals stopped by other officers do not end up in jail. That is, to her, other officers are less efficient than her in convicting criminals, and she assumes that they know this. This implies that she *perceives* that other officers have lower marginal revenue of an additional stop than she does. Since blacks are being oversampled (compared to strict random sampling) due to their higher crime rate, the cost can be reduced by stopping less blacks at the margin. Thus, she *imagines* that other officers are indeed stopping fewer blacks than she is. The jail data causes her to overestimate the black crime rate. To clarify, consider the following example. Suppose, in actuality, 10 blacks and 10 whites are stopped, and as a result, 2 blacks and 2 whites land up in jail. That is, the crime rates across blacks and whites are actually 20%. Now if the officer, using the above argument, believes the jail rates are a consequence of 8 blacks and 12 whites being stopped, she will estimate the black (white) crime rate to be 25% (16.7%). This last effect increases when the level of overconfidence increases.

In our model blacks are stopped more frequently than they would if fully rational police officers were practicing statistical discrimination. This result is consistent with the findings in Hernández-Murillo and Knowles (2004), where it is shown that there is in fact an excess sampling of minorities, above and beyond what can be justified on efficiency grounds.<sup>4</sup> This excess sampling of minorities can be explained by racism on the part of the police; this paper offers an alternative explanation. In fact, an interesting implication of the results we obtain is that over-incarceration of a specific ethnic group can occur, even if officers are not racist in the least. This aspect is important because substantial resources have been and are now being dedicated to ensuring that individual police officer do not harbor racism. As such it is of importance to assess whether this approach alone necessarily solves the problem of discrimination. We show that as long as officers are provided with incentives to catch as many criminals as possible, over-incarceration is purely a result of optimizing behavior on the part of non-racist police. Higher overconfidence can also worsen the racial imbalance in jails. Our model, which features a fully dynamic evolution of policer officers' perception of racial criminality, also highlights the crucial role played by the incentives to catch criminals in the long run level of incarceration. More precisely, although racial jail imbalance is a natural consequence of our statistical discrimination model, we show that it is worsened when officers become more overconfident.

---

<sup>4</sup>This work extends a paper by Knowles et al. (2001), where the opposite result was obtained, mainly due to the fact that their data did not discriminate between low and high discretion interdiction.

Our paper is part of a burgeoning literature in economics that explores the consequences of departures from Bayesian rationality on belief formation. Many of these contributions attempt to explain the causes overconfidence in human behavior. For instance, Rabin and Schrag (1999) offers an early model of an agent who is subject to confirmatory bias.<sup>5</sup> They show, in particular, that this agent exhibits overconfident behavior. Bénabou and Tirole (2002) develop a model where overconfidence is the result of endogenously chosen selective recall. More recently, Van den Steen (2004) shows that overconfidence can be an equilibrium behavior for strictly rational agents if they hold differing priors. Closer to our setup, Brocas and Carillo (2005) presents a model in which relative overconfidence arises naturally in a setup where agents learn about their own ability.<sup>6</sup> A closely related strand of this literature takes rational biases as a starting point and attempts to describe their economic and institutional consequences. Both Barber and Odean (2001) and Biais *et.al* (2005) study traders who have overconfidence in their own judgements and their impact on financial markets. They both find that overconfident traders trade more and have a reduced trading performance. Similarly, Landier and Thesmar (2005) study the features of financial contracts passed between investors and entrepreneurs and show that they result from the overoptimistic nature of the entrepreneurs. To the best of our knowledge, our paper is the first to connect a cognitive bias with discriminatory practices.

While our model shares certain aspects of belief formation and self-fulfilling negative discrimination with many other papers, such as Farmer and Terrel (1996), Acemoglu (1995), Coate and Loury (1993), and more recently, Verdier and Zenou (2001), the similarities are somewhat superficial. These models possess the common feature that there is a principal (often an employer) who acts upon a negative belief and takes a discriminatory action against the agent. This action by the principal, in turn, makes it optimal for the party who is discriminated against to take such actions as to make the principal's "prophecy" self-fulfilling. In our setup, when the current generation of police use incarceration data to decide on which race to interdict more frequently, they may perpetuate their mistaken belief against the over-represented race, but at a cost: the cost of deviating from strictly random sampling. The aforementioned papers do not model this cost. Two additional points of contrast deserve mention here. First, we clearly demonstrate that long-run discrimination against a race can be obtained even without the assumption of a priori negative

---

<sup>5</sup>A person who suffers from confirmatory bias tend to interpret ambiguous information as evidence confirming her current hypothesis.

<sup>6</sup>Zabojnik (2004) and Santos-Pinto and Sobel (2004) obtain results with similar flavor.

beliefs. That is, the very existence of a common and well documented rational bias is sufficient to generate a long-run discriminatory outcome. Second, because we do not discuss issues relating to the supply of crime, we are, per force, silent on whether the race that is discriminated against truly ends up becoming more criminal, thereby “justifying” the discrimination.<sup>7</sup>

## 2. A MODEL OF UPDATING WITH OVERCONFIDENCE

**2.1. The environment.** Consider a situation where at each date  $t = 1, 2, \dots, \infty$ , a fraction  $\gamma$  of the population is black and the remaining fraction is white. Let  $c_i$ ,  $i = b, w$  be the unknown, time-invariant, *true* percentage of criminals among the black and white populations, respectively. Note that since the number of convictions are influenced by the behavior of the police and self-reported crime rates are notoriously unreliable, these numbers are unknown in practice. Without loss of generality, we will assume throughout that  $c_b > c_w$ .<sup>8, 9</sup>

We model highway stops as follows: A police officer can legally stop any car, and subsequently determine with 100% accuracy whether the occupant is guilty of a serious offense, like drug smuggling. In that case, the person is always convicted and spends the remainder of the period in jail. Alternatively, if stopped, an innocent person is immediately released and not stopped for the rest of the period. We assume that the officer receives a compensation of  $\$x$  for each guilty person he catches, and nothing otherwise. The  $\$x$  is a proxy for all incentive rewards, such as pay increases, promotions, medals, etc.

The choice to model a highway stop is driven by two factors: the majority of data on racial profiling has been collected in the context of highway stops and the more discretion officers have, the more pronounced the racial profiling becomes.<sup>10</sup> This is demonstrated by

---

<sup>7</sup>Persico (2002) discusses the “general strain theory” and how unfairness in police practices may itself encourage crime among those being unfairly treated.

<sup>8</sup>Note that here we deviate from Persico (2002), who assumes that in equilibrium  $c_b = c_w$ . This is the natural outcome in his model, since he assumes that there is no cost associated with stopping only members of one race. This assumption implies that if  $c_b > c_w$ , only blacks will be stopped, which, in turn, will lead to the crime rate of whites to rise. When stopping only members of one race is costly to the officer, this may not be the equilibrium outcome.

<sup>9</sup>Our model is not meant to capture an important part of the work routine of highway officers, namely that of DWI or DUI offenses. There are at least two reasons for this. First, most DWI offenders do not end up in jail. Second, it is unclear whether the assumption  $c_b > c_w$  is appropriate when one considers these crimes, since most DWI offenses are committed by white drivers. We thank an anonymous referee for pointing out these points.

<sup>10</sup>A study by John Lamberth shows that 98.1% of all cars on a stretch of the New Jersey Turnpike are exceeding the speed limits. (See *New Jersey v. Soto*, 1996) This makes it reasonable to assume that police can stop any car passing by. Typically these violations are used as cause to stop the vehicle, but after the

an example in Verniero and Zoubek (1999): “[...] the Public Defender’s statistical expert compare tickets issued [...] by three different State Police Units: (1) the Radar Unit, which [...] exercises comparatively little discretion; (2) the Tactical Patrol Unit, which [...] exercises somewhat greater discretion; and, (3) the Patrol Unit, which [...] exercises the most discretion. [...] the Radar Unit was found to have issued 18% of its tickets to African Americans, the Tactical Patrol Unit issued 23.8% of its tickets to African Americans, and the Patrol Unit issued 34.2% of its tickets to African Americans.”

The officer makes a predetermined number  $n$  of highway stops each period. If the officer expends no special effort, he stops a racial mix which corresponds to the population proportion, that is  $\gamma n$  blacks and  $(1 - \gamma)n$  whites. At a cost, however, he can change these proportions. Let  $S_b^t$  denote the proportion of the people he stops in period  $t$  who is black.<sup>11</sup> The associated cost function is  $C(\gamma, S_b^t) = \frac{\gamma}{2} - S_b^t \left(1 - \frac{1}{2\gamma} S_b^t\right)$ . This is the cost per stop of deviating from strictly random sampling. Note that  $C(\cdot)$  obtains its minimum at  $S_b^t = \gamma$ , is convex, and symmetric around  $S_b^t = \gamma$ .<sup>12</sup>

A discussion of the cost function is in order. We assume that the cost function obtains its minimum at  $\gamma$  because random stopping is easier for the police officer, since there is no need to establish the race of the motorist before pulling him over, nor is there any need to wait for a driver of the right ethnic group. An example of the effort associated with determining the race of the driver can be found in the following quote in Verniero and Zoubek (1999): “[...] some troopers positioned their vehicles perpendicular to the roadway at night to spotlight the occupants of moving vehicles. [...] this latter practice (which renders radar guns less effective or useless) would seem to support the suspicion that officers had taken race and ethnicity into account.” At a deeper level, our cost function tries to capture the inherent feeling of outrage many people share regarding the use of race as an instrument of statistical discrimination.<sup>13</sup> Since many of these people ultimately help elect sheriffs and other police personnel, the average police officer on the street is indirectly

---

vehicle has been stopped it is searched for drugs and weapons or other signs of more serious offenses.

<sup>11</sup>In a variant of this model,  $S$  could be the portion of work day that the police officer devotes to patrol one neighborhood rather than another. Similarly to the racial fraction, it can be argued that the police officer has some discretion in the choice of  $S$ . There is again a moral hazard component here, in the sense that it is difficult to monitor accurately where and for what exact reason the police officer is in a given neighborhood.

<sup>12</sup>The symmetry of the cost function implies that it is equally costly to discriminate against either race.

<sup>13</sup>Note that we implicitly assume that  $(c_b - c_w)$  does not affect the cost function. It is plausible that society is more tolerant of unequal treatment when  $(c_b - c_w)$  is large. This could be incorporated into our model by assuming that the cost depends directly on  $(c_b - c_w)$  in a decreasing fashion. If this effect is strong enough it has the potential to overturn our results.

forced to recognize this outrage in her own decision-making.<sup>14</sup>

Before moving on, it is useful to formally define our notion of the fair steady state. This is one where the long run racial composition of the prison population is identical the racial composition of the criminal population, in and out of jail. The proportion of inmates who is black at the fair steady state is therefore  $F^* = \frac{c_b \gamma}{c_b \gamma + c_w (1 - \gamma)}$ . This provides a utopian benchmark to which we compare various long run outcomes.<sup>15</sup> This is also the steady state which would obtain in the model if we disallowed incentives.

## 2.2. The officer's optimization problem.

**Full information, no overconfidence.** Initially we consider, as a benchmark, the case where  $c_b$  and  $c_w$  are known and where there is no overconfidence. The officer's period  $t$  expected per stop payoff is:

$$R(S_b^t) = S_b^t \cdot c_b \cdot x + (1 - S_b^t) \cdot c_w \cdot x - \left( \frac{\gamma}{2} - S_b^t \left( 1 - \frac{1}{2\gamma} S_b^t \right) \right). \quad (1)$$

If  $c_b = c_w$  the optimal choice is clearly  $S_b^* = \gamma$ ; in this case, the racial distribution of the incarcerated population will correspond to the racial distribution of the criminal population in general. If  $c_b \neq c_w$ , the optimal  $S_b^*$  is easily seen to be  $S_b^* = \gamma + x\gamma(c_b - c_w)$ . This suggests that if blacks in fact engage in more criminal activity than whites, more than  $\gamma n$  of the people the officer stops will be black. Using  $S_b^*$ , we can calculate the steady state number of white and black inmates, denoted by  $P_b^{FI}$  and  $P_w^{FI}$  respectively. Each officer stops  $S_b^* n$  ( $(1 - S_b^*) n$ ) blacks (whites), and incarcerates  $c_b S_b^* n$  ( $c_w (1 - S_b^*) n$ ) black (white) criminals. With a total of  $m$  officers, the steady state jail populations are  $P_b^{FI} = c_b S_b^* nm$  and  $P_w^{FI} = c_w (1 - S_b^*) nm$ . The steady state proportion of inmates who is black is given by

$$J_b^{FI} \equiv \frac{P_b^{FI}}{P_b^{FI} + P_w^{FI}} = \frac{\gamma c_b (1 + x(c_b - c_w))}{c_b \gamma + c_w (1 - \gamma) + x\gamma(c_b - c_w)^2}.$$

<sup>14</sup>Four officers in Oklahoma were named in a suit because they had stopped relatively more blacks than the population proportions of the state. The suit was dropped when the statisticians on the defence team demonstrated that many drivers on the particular highway were in fact driving with Texas plates, and since Texas has a larger proportion of blacks, the stop data could be the result of random sampling.

<sup>15</sup>Katherine Kersten implies a similar definition when she argues that "... to determine whether Minnesota has "too many" male inmates, we must compare the proportion of males in the prison population with the proportion of males in the *criminal* population, not the population at large..." in her 2001 article in *The Weekly Standard*. Persico (2002) employs the definitions that both groups should be investigated with the same intensity, which in our setup would lead to a fair outcome. Gary S. Becker provides a somewhat different definition in his July, 2000 article in *Business Week* "... if stops of blacks and whites uncover evidence at about the same rate, that suggests the police are using reasonable criteria for deciding whom to stop and search." This will never occur in our model unless crime rates are identical, and even then this may not be the best definition of fairness. See Chakravarty (2002) for a detailed discussion on this.

It is easy to verify that  $J_b^{FI} > F^*$  if  $x > 0$ . Given the incentives provided to the officer, it is optimal for her to stop more of the higher crime race than she would under random sampling. She continues to stop some whites, because the cost of stopping only blacks exceeds its benefit. In the absence of incentives ( $x = 0$ ), the full information steady state is identical to the fair steady state. At the fair steady state the total number of inmates is  $(\gamma c_b + (1 - \gamma) c_w) nm$ , while at the full information steady state the number of inmates is strictly greater, namely  $P_b^{FI} + P_w^{FI} = \left( c_b \gamma + c_w (1 - \gamma) + x \gamma (c_b - c_w)^2 \right) nm$ . This difference, then, is purely due to the introduction of incentives to catch criminals.

It is worth noting that Persico (2002) and other papers obtain the fair steady state even though there is no cost to non-random sampling. In these papers, changes in the supply of crime cause crime rates to eventually equalize rendering random sampling optimal; consequently, the fair steady state is achieved. We find that when the crime rates *cannot* equalize, as is likely the case in the US, it is not possible to obtain the fair steady state when officers are provided with incentives to arrest criminals.

**2.3. Full information, overconfidence.** Now consider the slightly more involved case where the atomistic officer can accurately estimate  $c_b$  and  $c_w$ , but suffers from relative overconfidence. The relative overconfidence expresses itself in the mistaken belief that other officers only manage to incarcerate a fraction  $s$  of the criminals (black or white) that they encounter. The officer assumes that *other* officers are aware that their incarceration rate is  $s$  and they take this into account when they optimize. Note that we are assuming that the incarceration rates is identical for blacks and whites.<sup>16</sup>

The ability to accurately estimate the crime rates mimics a situation where (as currently mandated by law in many states) detailed data on the racial composition of stopped motorists was collected for stops, searches, seizures, and arrest, and made available to individual police officers. In the absence of overconfidence, this information would enable the officers to obtain consistent (and very low variance) estimates of the true crime rates. In effect, they would know the true crime rate, and as a result, the full information steady state would obtain. However, overconfident officers will still misinterpret the data, and estimate the crime rates using  $\hat{c}_b = \frac{1}{s} c_b$  and  $\hat{c}_w = \frac{1}{s} c_w$ , since she will assume that only a fraction  $s$  of the criminals are convicted. In that case, the steady state proportion of blacks

---

<sup>16</sup> Alternative plausible assumptions could be: "Whites are more cerebral, and hence harder to arrest" or "Blacks are more violent and hence harder to arrest" or "Whites have better lawyers, and are therefore harder to convict".

in jail is

$$J_{FIO}^* = \frac{\gamma c_b + \frac{x\gamma}{s} c_b (c_b - c_w)}{\gamma c_b + (1 - \gamma) c_w + \frac{x\gamma}{s} (c_b - c_w)^2}. \quad (2)$$

The next proposition compares this steady state to that obtained in the previous section. This result may be interpreted as describing the effect of overconfidence on jail rates.

**Proposition 1.** :

- a) *The FIO steady state always has a greater fraction of inmates who is black than the full information steady state (and hence greater than  $F^*$ ).*
- b) *Increasing incentives ( $x$ ) or higher overconfidence (lower  $s$ ) will increase the fraction of black jail inmates.*
- c) *FIO is identical to the fair steady state when there are no incentives to catch criminals ( $x = 0$ ).*
- d) *The total number of criminals caught under FIO exceeds the total number of criminals caught under the full information steady state.*

The proposition states that overconfidence increases the problem of over-incarceration of blacks.

**2.4. Unknown crime rates, overconfidence.** Now consider the more general case where the atomistic officer does not know  $c_b$  and  $c_w$  and suffers from relative overconfidence. We assume that the officer forms an estimate of these probabilities ( $\hat{c}_b$  and  $\hat{c}_w$  respectively) using published jail data.<sup>17</sup> Based on the estimates formed from period  $t - 1$  jail rates, the officer stops a fraction  $S_b^t = \gamma + x\gamma (\hat{c}_b^{t-1} - \hat{c}_w^{t-1})$  blacks in period  $t$ . Now consider an officer with incarceration rate  $s$ . A fraction  $\tilde{S}_b^t$  of the drivers she stops will be black; then  $\tilde{S}_b^t = \gamma + sx\gamma (\hat{c}_b^{t-1} - \hat{c}_w^{t-1})$ . For future use, we assume  $s > \bar{s} \equiv \max\{\frac{-1 + \sqrt{1 + 8c_b}}{2}, \frac{-(1 - \gamma) + \sqrt{(1 - \gamma)^2 + 4\gamma c_w}}{2\gamma}\}$  (note that  $\bar{s} < 1$ ) and  $\frac{1}{2} \frac{1}{x(c_b - c_w) + 1} < \gamma < \frac{1}{x + 1}$ . The former condition guarantees the existence of an interior steady state and the latter ensures that  $S_b^t$  is a strictly positive fraction.

In any period  $t$ , there are  $P_b^t$  blacks and  $P_w^t$  whites in jail. Since the officer is atomistic,

---

<sup>17</sup>The reasons for using jail data as opposed to only her own data could be many. For example the fact that jail data provides significantly more observations or that local crime rates fluctuate due to movement of criminals. Mostly, however, we consider this assumption a proxy for the fact that officers' beliefs are influenced by briefings, discussions with their colleagues etc.

she believes that these jail numbers are created as follows:

$$\begin{aligned} P_b^t &= s c_b \tilde{S}_b^{t-1} n m, \\ P_w^t &= s c_w \left(1 - \tilde{S}_b^{t-1}\right) n m. \end{aligned}$$

Note that she perceives  $\tilde{S}_b^{t-1}$  to be the fraction of stopped drivers who is black. Thus, her estimate of  $c_b$  should satisfy

$$P_b^t = s \hat{c}_b^t \tilde{S}_b^{t-1} n m, \quad (3)$$

and similarly

$$P_w^t = s \hat{c}_w^t \left(1 - \tilde{S}_b^{t-1}\right) n m. \quad (4)$$

Notice that her estimates of criminality among the races is tainted by her overconfidence directly through  $s$  and indirectly through  $\tilde{S}_b$ . Clearly, since all the officers in reality have an incarceration rate of 1, the jail numbers are *actually* generated in the following manner:

$$P_b^t = c_b S_b^{t-1} n m, \quad (5)$$

$$P_w^t = c_w \left(1 - S_b^{t-1}\right) n m \quad (6)$$

Thus, at date  $t$ , the estimates  $\hat{c}_b^t$  and  $\hat{c}_w^t$  are generated by the following two equations:

$$\hat{c}_b^t = \frac{c_b S_b^{t-1}}{s \tilde{S}_b^{t-1}} \quad (7)$$

and

$$\hat{c}_w^t = \frac{c_w \left(1 - S_b^{t-1}\right)}{s \left(1 - \tilde{S}_b^{t-1}\right)}. \quad (8)$$

Now, define  $\hat{k}^t \equiv \hat{c}_b^t - \hat{c}_w^t$ , and plug the expression for  $\tilde{S}_b^{t-1}$  into (7) and (8). We can then simplify the system to the following non-linear, first-order difference equation:<sup>18</sup>

$$s \hat{k}^t = \frac{c_b \left(1 + x \hat{k}^{t-1}\right)}{1 + s x \hat{k}^{t-1}} - \frac{c_w \left(1 - \gamma - x \gamma \hat{k}^{t-1}\right)}{1 - \gamma - s x \gamma \hat{k}^{t-1}}, \quad t = 1, 2, \dots \quad (9)$$

Define  $k^*$  as the steady state value of  $\hat{k}^t$  and  $J_b^*$  as the steady state value of the proportion of blacks in jail,  $\frac{P_b^*}{P_b^* + P_w^*}$ . Our main result can be stated as follows:

---

<sup>18</sup>At time 1, we will assume that there is an existing jail population,  $P_b^0$  and  $P_w^0$ . The model is thus initiated by the officer calculating her first estimates, maximizing revenue and determining the optimal number of blacks (and whites) to stop. It turns out that the composition of the initial jail population does not affect the long run jail rates.

**Proposition 2. :**

- a) *There exists a unique, interior, steady state  $k^*$ , such that  $0 < k^* < 1$ .*
- b) *The steady state proportion of blacks in jail ( $J_b^*$ ) is greater than the corresponding proportion under full information with overconfidence ( $J_b^{FIO}$ ), and hence greater than the fair rate ( $F^*$ ).*
- c) *As the reward for arrests ( $x$ ) increases, the steady state proportion of blacks in jail ( $J_b^*$ ) increases.*
- d) *If there are no incentives ( $x = 0$ ), the fair steady state is obtained.*
- e) *The total number of criminals caught under the steady state with overconfidence exceeds the total number of criminals caught under the FIO steady state.*

Proposition 2 states that in the long run, blacks will be overrepresented in the jail population, more than what can be justified by their assumed higher crime rate. The intuition for this result is as follows: An officer's overconfidence bias causes her to believe that other officers manage to incarcerate only a fraction of the criminals they stop. Thus she perceives that the marginal benefit of each of their stops is lower than her own. Payoff maximization would entail that they reduce their marginal cost; which they can do by stopping less blacks at the margin, since that would reduce the cost of deviating from random sampling. Consequently, she underestimates the number of black drivers stopped, and overestimates the black crime rate implied by the jail data.

As the incentives increase the overincarceration of blacks worsens. As  $x$  increases, marginal revenue increases, but the marginal cost is left unchanged. At the optimum, then, more blacks can be stopped. Similarly, when all incentives are removed, revenue is maximized when the police practice random sampling; the fair steady state results.

The overestimated black crime rate relative to the white causes more blacks to be stopped than would have been the case under full information. Since blacks are assumed to have a higher crime rate, the result is a higher jail population overall.<sup>19</sup> Thus, our result points to a clear trade-off between fairness and efficiency, an issue we take up in Bunzel and Marcoul (2005).

Before proceeding further, it is important to touch on two items, the stability and certain comparative static properties of the steady state. Table 1 presents results of simu-

---

<sup>19</sup>This result contrasts with the results in Persico (2002), where he finds that if both races have the same population fraction of criminals (an equilibrium free-entry-into-crime condition in his static model) and if one group is policed with less intensity, then forcing the police to increase the intensity of interdiction on that group could raise both fairness and effectiveness.

lation exercises which suggest that for a wide range of parameters,  $k^*$  is locally stable. How does  $k^*$  respond to an increase in the overconfidence parameter,  $s$ ? Figures 1-3 confirm our intuition: as the level of overconfidence increases, the excess jailing of blacks increases as well. Figures 4 and 5 depict the steady state as a function of  $x$ .

### 3. CONCLUSION

We develop a stylized dynamic model of highway policing in which a non-racist police officer exhibits relative overconfidence. The officer is given incentives to arrest criminals but faces a per stop cost which increases when the racial mix of her stops differs from that of the population. Each period, she observes the racial composition of jails and forms estimates about the crime rates of each race. In some settings, her overconfidence leads her to overestimate the crime rate of one race relative to another causing the long-run racial composition of the jail population to worsen when compared to the racial composition which would obtain without overconfidence. We compare the model where overconfident officers estimate crime rates to a situation where officers have detailed stop data for each race, similar to data being currently collected in many US states. While such data collection is shown to go a long way to ameliorate the problem of over incarceration of blacks, the cognitive bias of the officers (and their incentives) prevent a complete elimination of the problem. This policy prescription is in direct conflict with the current movement to use voter initiatives to ban the registration of race in society. The de-facto leader of this movement, Ward Connerly, argues that racial data collection is not only costly for taxpayers but more importantly “stigmatize” minorities even further (blacks in particular). He argues that by focussing public attention on minorities, this type of data collection worsens the extent of racial discrimination and that all people have a right to “racial privacy.” The upshot of this line of argument is that the race of an individual is not important and should not be a matter of public policy. Our position on that subject is that racial data collection is desirable for the reasons stressed above. Government should not be uninformed about discrimination practices.<sup>20</sup>

An issue which should be addressed in future research is the paradox that although the rights of the black population have never been so well guaranteed as they are today, the racial composition of US jails has never appeared as much in their disfavor. Indeed, African-

---

<sup>20</sup>It is worth noting that other types of biases could provide results which differ from those obtained here. In particular, as pointed out by a referee, if all officers believe that they are more ethical (less racist) than their colleagues, this would lead to *underestimation* of the black crime rate, which in turn would lead to increased fairness in equilibrium.

Americans represented about 52% of state prison inmates in 1991, up from 34% in 1960 and 23% in 1926. There is no consensus on the cause of this trend, though Harcourt (2004) suggests that it has occurred mostly because of a dramatic evolution in the investigatory search and seizure jurisprudence that has confirmed and reinforced the rights given to police to use traffic violations as a valid pretext for stop-and-frisk procedures when they suspect other more severe violations, such as drug trafficking. Such arguments are best studied in dynamic models of law enforcement.

Another issue which deserves attention in the broader literature on racial profiling is the trade off between resources spent on stop-and-frisk procedures, which in essence are a pre-emptive policing technique aimed at avoiding the most disastrous consequences of crime, and resources spent on “ex post” policing, i.e., after the crime has been committed. We leave these issues for future research.

## REFERENCES

- [1] Acemoglu, D. (1995), "Public policy in a model of long-term unemployed", *Economica* 62, 161-178.
- [2] American Civil Liberties Union Freedom Network (1999), "Driving While Black. Racial Profiling on Our Nation's Highways." *Special Report*, June.
- [3] Babcock, L. and G. Lowenstein (1997), "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives*, v11, n1, 109-26.
- [4] Barber, B. and Odean, T. (2001) "Boys will be boys: Gender, Overconfidence and common stock Investments," *Quarterly Journal of Economics*, Vol. 116 (3), pp. 261-292.
- [5] Becker, G. S. (1971), *The Economics of Discrimination*, University of Chicago Press, Chicago.
- [6] — (2000), "Tough Justice Is Saving Our Inner Cities," *Business Week*, July 17.
- [7] Bénabou, R. and Tirole, J. (2002) "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics*, Vol. 117 (3), pp. 871-915.
- [8] Brocas, I. and Carrillo, J. (2005) "Biased Decision-Making by Agents with Unbiased Beliefs," Mimeo, University of Southern California.
- [9] Bunzel, H., and P. Marcoul. (2005), "On the Use of Racial Profiling as a Law Enforcement Tool", Working Paper, Iowa State University.
- [10] Borooah, V.K. (2001), "Racial bias in police stops and searches: an economic analysis" *European Journal of Political Economy* 17, 17-37.
- [11] — (2002), "Racial bias in police stops and searches: a reply," *European Journal of Political Economy* 18, \ 607-608.
- [12] Chakravarty, S.P. (2002), "Economic analysis of police stops and searches: a critique," *European Journal of Political Economy* 18, 597-605.
- [13] Coate, S. and G.C. Loury (1993), "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review* 83, 1220-1240.
- [14] Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998), "Investor psychology and security market under- and overreactions", *Journal of Finance* 53, 1839-1885.
- [15] DePaulo, B. M. and R. L. Pfeifer (1986) "On-the-job experience and skill at detecting deception and truth," *Journal of Applied Social Psychology*, 16(3), 249-267.
- [16] Doster, S. (2001), "Racial Profiling Law Lacks Teeth; Loophole Lets Police Off Hook, Some Say," *The Times-Picayune* (New Orleans), December 16.
- [17] Dunphy, J. (2001), "Counting Noses," *National Review*, December 21.
- [18] Farmer, A. and D. Terrel (1996), "Discrimination, Bayesian updating of employers belief, and human capital accumulation", *Economic Inquiry* 34, 204-219.

- [19] Garrido, E., J. Masip and C. Herrero (2004, forthcoming) "Police officers' credibility judgements: Accuracy and estimated ability," *International Journal of Psychology*.
- [20] Glaser, M., and M. Weber, 2005, "Overconfidence and trading volume," Mimeo, university of manheim.
- [21] Harcourt, B. (2004). "The shaping of Chance: Actuarial Models and Criminal Profiling at the turn of the Twenty-First Century", *The university of Chicago Law Review*, 70: 105-128.
- [22] Hernández-Murillo, R. and Knowles, J. (2004) "Racial profiling or racist policing? Bounds tests in aggregate data.", *International Economic Review*, 45(3), 959-989.
- [23] Kahneman, D. and Riepe, M. W. (1998), "Aspects of investor psychology", *Journal of Portfolio Management*, 24(4), 52-65.
- [24] Kersten, Katherine (2001), "Race to Conclusions: What the activists don't tell you about racial profiling," *The Weekly Standard*, August 20.
- [25] Knowles J., P. Todd , and N. Persico. (2001), "Racial Bias in Motor-Vehicle Searches: Theory and Evidence," *Journal of Political Economy* 109(1), . 203-29.
- [26] Landier, A. and Thesmar, D. (2005) "Optimistic Beliefs As Private Benefits For Entrepreneurs: Implication for Contracting and evidence", mimeo.
- [27] Langan, A.P., L. A. Greenfeld, S. K. Smith, D. J. Levin and M. Durose "Contacts between Police and the Public-Findings from the 1999 National Survey", Bureau of Justice Statistics, NCJ-184957.
- [28] Meissner, C.A. and S. M. Kassin (2002) " "He's guilty!": Investigator bias in judgments of truth and deception," *Law and Human Behavior*, 26(5), 469-480.
- [29] Meyer, H. (1975), "The pay-for-performance dilemma," *Organizational Dynamics*, 3, 39-50.
- [30] Moro, A. and P. Norman (2004) "A general equilibrium model of statistical discrimination," *Journal of Economic Theory*, 114(1), 1-30
- [31] Mullin, R. (1995) "Police Judgement", *Policing* v11(4)
- [32] Nöth, M. and Weber, M. (2003), "Information aggregation with random ordering: Cascades and overconfidence", *Economic Journal* 113(484), 166-189
- [33] Persico N. (2002), "Racial Profiling, Fairness, and Effectiveness of Policing," *The American Economic Review*, 92(5), 1472-97.
- [34] Rabin, M. and Schrag, J. (1999) "First Impression Matters: A model of Confirmatory Bias." *Quarterly Journal of Economics*, Vol 114 (1). pp. 37-82.
- [35] Santos-Pinto, L. and J. Sobel (2004), "A Model of Positive Self-Image in Subjective Assessments", forthcoming in *American Economic Review*.

- [36] Svenson, O. (1981), "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica*, 47, 143-148.
- [37] Van den Steen, E. (2004), "Rational Overoptimism (and Other Biases)," *American Economic Review*, Vol. 94, pp. 1141-1151
- [38] Verdier T. and Y. Zenou. (2001), "Racial Beliefs, Location and the Causes of Crime," mimeo.
- [39] Verniero, P and P.H Zoubeck "Interim Report of the State Police Review Team Regarding Allegations of Racial Profiling", April 20, 1999.
- [40] Yates, J. F. (1990), *Judgment and Decision Making*, Prentice-Hall, Englewood Cliffs.
- [41] Zabojsnik, J. (2004), "A Model of Rational Bias in Self-Assessment", *Economic Theory*, Vol. 23, 259-282.

## Appendix

### A. PROOF OF PROPOSITION 1

**Proof of a)**  $J_{FIO}^*$  is greater than the full information steady state if

$$J_{FIO}^* = \frac{\gamma c_b \left(1 + \frac{1}{s} x (c_b - c_w)\right)}{\gamma c_b + (1 - \gamma) c_w + \frac{\gamma x}{s} (c_b - c_w)^2} > \frac{\gamma c_b (1 + x (c_b - c_w))}{c_b \gamma + c_w (1 - \gamma) + x \gamma (c_b - c_w)^2} = J_b^{FI} \Leftrightarrow$$

$$1 > s,$$

which holds by assumption.

**Proof of b)** We need to show that  $\frac{\partial J_{FIO}^*}{\partial x} > 0$  and  $\frac{\partial J_{FIO}^*}{\partial s} < 0$ . Simple differentiation provides

$$\frac{\partial J_{FIO}^*}{\partial x} = \frac{\gamma c_b c_w (c_b - c_w) s}{\left[\gamma c_b + (1 - \gamma) c_w - \frac{x \gamma}{s} (c_b - c_w)^2\right]^2} > 0,$$

$$\frac{\partial J_{FIO}^*}{\partial s} = -\frac{\gamma x c_b c_w (c_b - c_w)}{\left[\gamma c_b + (1 - \gamma) c_w - \frac{x \gamma}{s} (c_b - c_w)^2\right]^2} < 0.$$

**Proof of c)** Simply set  $x = 0$  in the expression of  $J_{FIO}^*$  and the result is evident.

**Proof of d)** The total number of criminals caught under FIO is

$$P_b^{FIO} + P_w^{FIO} = \left(\gamma c_b + (1 - \gamma) c_w + \frac{\gamma x}{s} (c_b - c_w)^2\right) nm.$$

The total number of criminals caught under the full information steady state is

$$P_b^{FI} + P_w^{FI} = \left(c_b \gamma + c_w (1 - \gamma) + x \gamma (c_b - c_w)^2\right) nm., \text{ so}$$

$$P_b^{FIO} + P_w^{FIO} > P_b^{FI} + P_w^{FI} \Leftrightarrow$$

$$\gamma c_b + (1 - \gamma) c_w + \frac{\gamma x}{s} (c_b - c_w)^2 > c_b \gamma + c_w (1 - \gamma) + x \gamma (c_b - c_w)^2$$

$$1 > s,$$

which holds by definition of  $s$ .

### B. PROOF OF PROPOSITION 2

**Proof of a)** From (9), we know the steady state value of  $k$ ,  $k^*$  is defined by:

$$s k^* = \frac{c_b (1 + x k^*)}{1 + s x k^*} - \frac{c_w (1 - \gamma - x \gamma k^*)}{1 - \gamma - s x \gamma k^*}. \quad (10)$$

The task is to show that if  $\bar{s} \leq s \leq 1$ , then there exists a unique  $k^*$  in the interval  $(0; 1)$ . To this end we will need the following two lemmas:

**Lemma 3.** If  $s \geq y_1 = \frac{-1 + \sqrt{1 + 4(1+x)xc_b}}{2x}$ , there is at least one steady state in the interval  $(0; 1)$ .

**Lemma 4.** If  $s > y_2 = \frac{-(1-\gamma) + \sqrt{((1-\gamma)^2 + 4c_w\gamma x(1-\gamma+x\gamma))}}{2x\gamma}$  there is an even number of steady states in the interval  $[-1, 0]$ .

**Proof of Lemma 3:** We start by rewriting the steady state definition (10) in the following manner:

$$(1 + xsk)(1 - \gamma - \gamma xsk)sk = c_b(1 + xk)(1 - \gamma - \gamma xsk) - c_w(1 - \gamma - \gamma xk)(1 + xsk). \quad (11)$$

First notice that there is at most 3 steady states.

Define the left (right) hand side of (11) as  $LHS$  ( $RHS$ ). By plugging  $k = -1, 0, 1$  into the definitions of  $LHS$  and  $RHS$ , we obtain

$$LHS|_{k=-1} = -s(1 - xs)(1 - \gamma + x\gamma s) \quad (12)$$

$$RHS|_{k=-1} = c_b(1 - x)(1 - \gamma + x\gamma s) - c_w(1 - \gamma + x\gamma)(1 - xs) \quad (13)$$

$$LHS|_{k=0} = 0 \quad (14)$$

$$RHS|_{k=0} = (c_b - c_w)(1 - \gamma)$$

$$LHS|_{k=1} = (1 + xs)(1 - \gamma - x\gamma s)$$

$$RHS|_{k=1} = c_b(1 + x)(1 - \gamma - x\gamma s) - c_w(1 - \gamma - x\gamma)(1 + xs)$$

It is obvious from (14) that  $RHS|_{k=0} > LHS|_{k=0}$ . Therefore, for a single steady state to occur in the interval  $(0, 1)$ , it is necessary that the  $LHS$  and  $RHS$  cross in the interval. This occurs if  $RHS|_{k=1} < LHS|_{k=1}$ . The implied condition is

$$\frac{1+x}{1+xs}c_b - \frac{1-\gamma-x\gamma}{1-\gamma-x\gamma s}c_w < s.$$

It is possible to show that this is satisfied if and only if  $s > w$ , where  $w$  is implicitly defined by

$$(1+x)(1-\gamma-x\gamma w)c_b - (1-\gamma-x\gamma)(1+xw)c_w - w(1+xw)(1-\gamma-x\gamma w) = 0.$$

We will instead work with a sufficient condition which provides a somewhat neater condition, namely

$$\frac{1+x}{1+xs}c_b < s,$$

which can be written as

$$s \geq y_1 = \frac{-1 + \sqrt{1 + 4(1+x)xc_b}}{2x}. \quad (15)$$

It is easily seen that  $0 < y_1 < 1$ . Thus, for any  $s \in (y_1, 1)$ , (15) will hold with strict inequality.

**Proof of Lemma 3:** Since  $RHS_{k=0} > LHS|_{k=0}$ , there will be an even number of steady states between  $-1$  and  $0$  if  $RHS_{k=-1} > LHS|_{k=-1}$ . From (12) and (13) we obtain

$$c_b(1-x)(1-\gamma+x\gamma s) - c_w(1-\gamma+x\gamma)(1-xs) > -s(1-xs)(1-\gamma+x\gamma s) \Leftrightarrow \quad (16)$$

$$s > c_w \frac{(1-\gamma+x\gamma)}{(1-\gamma+x\gamma s)} - c_b \frac{(1-x)}{(1-xs)}.$$

Simple algebra verifies that a sufficient condition for this to hold is that

$$s > y_2 = \frac{-(1-\gamma) + \sqrt{\left((1-\gamma)^2 + 4c_w\gamma x(1-\gamma+x\gamma)\right)}}{2x\gamma}.$$

Note that  $y_2 < 1$ .

**Proof of Proposition 2:** First recall that there is at most 3 steady states. Now, if  $s \geq \max(y_1, y_2)$ , by Lemma 3, there is either 1 or 3 steady states between  $0$  and  $1$ , and we know from Lemma 4 that there is an even number of steady states between  $-1$  and  $0$ . We now have to determine whether there are 1 or 3 steady states between  $0$  and  $1$ . We will do this by determining whether we can document one or more steady states in the intervals  $(-\infty; -1)$  and  $(1; \infty)$ . To that end note that

$$\begin{aligned} \lim_{k \rightarrow \infty} LHS_{k \rightarrow \infty} &= -\infty, \\ \lim_{k \rightarrow -\infty} LHS_{k \rightarrow -\infty} &= +\infty, \\ \lim_{k \rightarrow \infty} RHS_{k \rightarrow \infty} &= -\infty, \\ \lim_{k \rightarrow -\infty} RHS_{k \rightarrow -\infty} &= -\infty. \end{aligned}$$

Recall that  $RHS_{k=-1} > LHS|_{k=-1}$ . Thus, since  $\lim_{k \rightarrow -\infty} RHS_{k \rightarrow -\infty} < \lim_{k \rightarrow -\infty} LHS_{k \rightarrow -\infty}$ , there must be at least one steady state in the interval  $(-\infty; -1)$ . This rules out the possibility of three steady states in the interval  $(0; 1)$ , and thus there must be only one. This leaves one steady state, but since there is an even number of steady states between  $-1$  and  $0$ , we can confirm that there are no steady states in this interval. Together with Lemmas 3 and 4, the conclusion is that there is exactly one steady state in each of the intervals  $(-\infty; -1)$  and  $(0; 1)$  and no steady state in the interval  $[-1; 0]$ .

**Proof of b)** Now turn the attention to the steady state between  $0$  and  $1$ . The steady state jail populations can be described by the following equations:

$$\begin{aligned} P_b^* &= c_b(\gamma + x\gamma k^*)nm \\ P_w^* &= c_w(1-\gamma - x\gamma k^*)nm \end{aligned}$$

These, in turn, provide the steady state jail rate for blacks:

$$J_b^* = \frac{P_b^*}{P_b^* + P_w^*} = \frac{\gamma c_b(1 + xk^*)}{c_b\gamma + c_w(1-\gamma) + (c_b - c_w)x\gamma k^*}.$$

Recall that the full information steady state jail rate for blacks is defined as  $J_b^{FIO} = \frac{\gamma c_b (1 + \frac{1}{s} x (c_b - c_w))}{\gamma c_b + (1 - \gamma) c_w + \frac{\gamma x}{s} (c_b - c_w)^2}$ . We thus need to find the conditions under which  $J_b^* > J_b^{FIO}$ .

$$\begin{aligned} J_b^* &> J_b^{FIO} \Leftrightarrow \\ \frac{\gamma c_b (1 + x k^*)}{c_b \gamma + c_w (1 - \gamma) + (c_b - c_w) x \gamma k^*} &> \frac{\gamma c_b (1 + \frac{1}{s} x (c_b - c_w))}{\gamma c_b + (1 - \gamma) c_w + \frac{\gamma x}{s} (c_b - c_w)^2} \Leftrightarrow \\ k^* &> \frac{1}{s} (c_b - c_w). \end{aligned} \quad (17)$$

Since  $0 < k^* < 1$ , the following inequality holds.

$$\frac{1 - S_b^*}{1 - \widetilde{S}_b^*} < 1 < \frac{S_b^*}{\widetilde{S}_b^*}.$$

It then follows from (10) that

$$k^* = \frac{c_b S_b^*}{s \widetilde{S}_b^*} - \frac{c_b (1 - S_b^*)}{s (1 - \widetilde{S}_b^*)} > \frac{1}{s} (c_b - c_w),$$

and hence, from (17), we know that  $J_b^* > J_b^{FIO}$ .

**Proof of c)** At date  $t \geq 1$ , the difference equation in  $k$  is given by

$$k_t = f(k_{t-1}) = \frac{c_b (1 + x k_{t-1})}{s (1 + s x k_{t-1})} - \frac{c_w (1 - \gamma - x \gamma k_{t-1})}{s (1 - \gamma - s x \gamma k_{t-1})}.$$

Note that

$$\frac{dk_t}{dk_{t-1}} = \frac{x(1-s)}{s} \left( \frac{c_b}{(1+s x k)^2} + \frac{c_w \gamma (1-\gamma)}{(1-\gamma-s x \gamma k)^2} \right) > 0.$$

It is routine to check that

$$\frac{d}{dx} \left( \frac{dk_t}{dk_{t-1}} \right) = \frac{(1-s)}{s} \left( c_w \frac{\gamma(1-\gamma)(1-\gamma+s x \gamma k)}{(1-\gamma-s x \gamma k)^3} + c_b \frac{1-s x k}{(1+s x k)^3} \right) > 0.$$

The steady state is defined as the intersection between the 45° degree line and  $f(\cdot)$ . Since  $k = f(0) = \frac{c_b - c_w}{s} > 0$ , when a positive steady state exists,  $f(k)$  will cross the 45° degree line from above. An increase in  $x$  will leave the curve unchanged at 0, but increase the slope everywhere. Hence the level of the steady state will increase (see Figure A1).

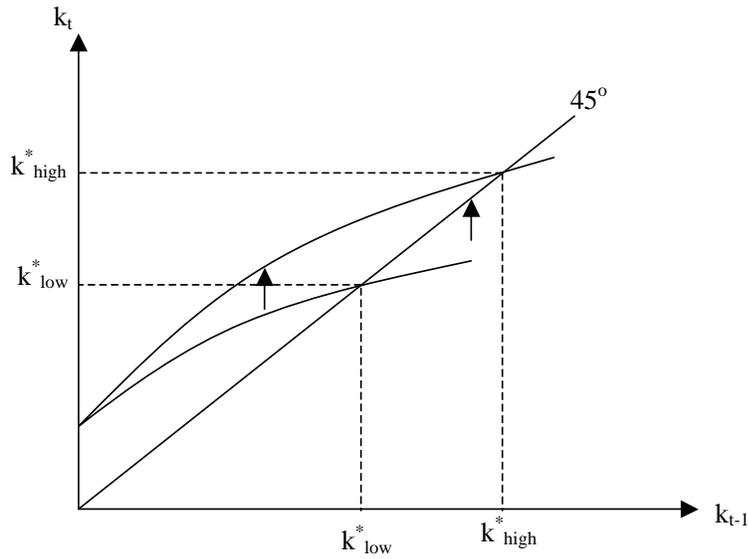


Figure A1: Effect on the steady state from an increase in  $x$

**Proof of d)** Simply plug  $x = 0$  into the expression of  $J_b^*$  to obtain the result.

**Proof of e)** The total number of criminals caught at the steady state is

$$P_b^* + P_w^* = (\gamma c_b + (1 - \gamma) c_w + (c_b - c_w) x \gamma k^*) nm.$$

Under full information with overconfidence, the total number of criminals caught is

$$P_b^{FIO} + P_w^{FIO} = \left( \gamma c_b + (1 - \gamma) c_w + \frac{\gamma x}{s} (c_b - c_w)^2 \right) nm.$$

Since, by (17),  $k^* > (c_b - c_w)$ , it is evident that  $P_b^* + P_w^* > P_b^{FIO} + P_w^{FIO}$ .

**Table 1:** Numerical results,  $c_b = 0.21$ ,  $c_w = 0.11$ 

$\gamma$	$x$	$s$	$k^*$	$\frac{\partial k_t}{\partial k_{t-1}}(k^*)$	$R(S_b(k^*))$
0.20	0.50	0.30	0.45	0.25	0.06
0.40	0.50	0.30	0.49	0.31	0.07
0.60	0.50	0.30	0.60	0.46	0.07
0.40	1.00	0.30	0.83	0.56	0.05
0.20	0.50	0.50	0.23	0.11	0.06
0.40	0.50	0.50	0.23	0.13	0.07
0.60	0.50	0.50	0.25	0.19	0.08
0.20	1.00	0.50	0.25	0.19	0.13
0.40	1.00	0.50	0.27	0.25	0.15
0.20	1.50	0.50	0.29	0.26	0.19
0.40	1.50	0.50	0.32	0.36	0.21
0.20	2.00	0.50	0.32	0.31	0.24
0.20	2.50	0.50	0.36	0.34	0.29
0.20	3.00	0.50	0.39	0.36	0.32
0.20	0.50	0.70	0.15	0.05	0.07
0.40	0.50	0.70	0.15	0.06	0.08
0.60	0.50	0.70	0.16	0.08	0.09
0.20	1.00	0.70	0.16	0.09	0.13
0.40	1.00	0.70	0.16	0.11	0.15
0.20	1.50	0.70	0.16	0.12	0.20
0.40	1.50	0.70	0.17	0.16	0.23
0.20	2.00	0.70	0.17	0.14	0.26
0.20	2.50	0.70	0.18	0.17	0.33
0.20	3.00	0.70	0.19	0.18	0.39
0.20	0.50	0.90	0.11	0.01	0.07
0.40	0.50	0.90	0.11	0.01	0.08
0.60	0.50	0.90	0.11	0.02	0.09
0.20	1.00	0.90	0.11	0.02	0.13
0.40	1.00	0.90	0.11	0.03	0.15
0.20	1.50	0.90	0.12	0.03	0.20
0.40	1.50	0.90	0.12	0.04	0.23
0.20	2.00	0.90	0.12	0.04	0.26
0.20	2.50	0.90	0.12	0.05	0.33
0.20	3.00	0.90	0.12	0.05	0.40