Spontaneous Discrimination

Marcin Pęski*and Balázs Szentes†

September 27, 2011

Abstract

This paper considers a dynamic economy where agents are repeatedly matched with one another and decide whether to form a profitable partnerships. Each agent has a physical colour and a *social colour*. The social colour of an agent is a signal about the physical colour of agents in his partenrship history. Before an agent makes a decision, he observes the other's physical and social colours. Neither the physical colour, nor the the social colour is payoff-relevant.

We identify environments where, in some equilibria, agents condition their decisions on the physical and social colours of their parners, that is, they discriminate. The main result of the paper is that every *stable* equilibria must involve discrimination in these environments. In particular, the colour-blind equilibrium is unstable.

1 Introduction

Consider a white community in the south where some people are members of the Ku Klux Klan. They dislike and are willing to punish anybody who is not white, and even those who are *associated* to non-whites. The rest of the community has no bias against people of other skin colour. People observe, perhaps imperfectly, each others' interactions in the community. Suppose now that a non-white individual seeks employment in this community. Of course, he will not be hired by Klan members. But even the unbiased people might be reluctant to give him a job because they are afraid of being punished by the Klan. At the end, the non-white individual could remain unemployed, that is, the whole community might end up discriminating against non-whites. Crucial in this story is that individuals observe some information about the interactions of others, for otherwise, unbiased individuals would not be afraid of hiring. But how large does the Klan have to be in order to sustain discrimination? This paper shows that even if nobody belongs to the Klan, the community might end up discriminating against non-whites.

This paper puts forward a new theory of racial discrimination. This theory is based on the assumption that each individual carries information about the physical colour of those to whom

*Department of Economics, University of Toronto, Toronto, CA.

†Department of Economics, London School of Economics, London, UK. E-mail: b.szentes@lse.ac.uk.

he is associated through past interactions. Individuals can condition their decisions on these information. The main result of this paper is that individuals might prefer to interact with others of their own colour and with those who are not associated to other colours. In other words, being associated to one's own physical colour becomes valuable through the equilibrium play.

To give an intuition for this result, let us revisit the story above and consider an unbiased member of the community. He might not want to interact with anybody who employed a non-white worker because he is afraid of being punished in the future for being indirectly associated to non-whites. In fact, for the same reason, he might not want to interact with those who are only indirectly associated to non-whites but never employed them. As a result, employing non-whites and being associated to them is punished not only by the Klan, but by the unbiased community members who are concerned about their associations. This concern can be self-enforcing and independent of the Klan. If unbiased individual are reluctant to interact with those who are associated with non-whites then unbiased individuals are better off not being associated to non-whites.

In the specific model analysed in this paper, agents are repeatedly matched with each other. After being matched, agents have to decide whether or not to enter into a profitable relationship. Each agent maximizes the discounted present value of expected monetary payoffs. Every relationship generates positive payoffs for both parties. Each agent has a physical colour which is either black or white. Before an agent makes a decision about forming a relationship, he observes the physical colour of his potential partner and an additional piece of information about his history of relationships. We model this information as a binary signal, either black or white, and refer to it as the *social colour* of the agent. If an agent decides to enter into a partnership with another one, his social colour becomes the physical or social colours of his partner with positive probability.

We characterize *stable* equilibria in our model. An equilibrium is called stable if, after perturbing the equilibrium strategies slightly, a myopic best-response dynamics converges back to the equilibrium. The main result of the paper is that each stable equilibrium involves discrimination under certain conditions. In particular, the colour-blind equilibrium, where agents ignore the physical and social colours, is unstable. In these cases, there are three stable equilibria. One in which the two races are segregated: members of each race discriminate against the other race. In the other two equilibria, discrimination is one-sided: one race discriminates against the other one, while the strategies of the members of the other race are colour-blind.

We emphasize that both the physical and social colours of the agents are payoff-irrelevant. Each agent cares only about monetary payoffs, and have no intrinsic preferences for the colours of his partners. Therefore, the equilibrium discrimination in our model is not *taste-based*. In addition, agents of different types are identical in terms of payoff-relevant characteristics both from the ex ante and the ex-post points of view. That is, the colours of the agents reveal no information regarding the profitability of his partnership. Therefore, the discrimination in our model is not a statistical one.

There is a large literature on taste-based discrimination, see Becker (1971) and Schelling (1971). These theories explain racial discrimination by assuming that individuals derive disutility from interacting with members of another race. Such preferences may be the result of group selection if a group does well relative to other groups if its members cooperate only with other group members. Alternatively, they might be outcomes of group formation processes. Similar people tend to have similar backgrounds which equips them with similar tastes, values, and attitudes. This facilitates making collective choices (Baccara and Yariv (2008), see also Alesina and Ferrara (2005)).

A common critique of taste based theories is that employers who do not discriminate make larger profits than those who discriminate and, hence, they would overtake competitive markets. In our model, an employer who does not discriminate also has higher instantaneous profits. However, these short-term gains from colour-blind hiring policy is offset by being boycotted by members of his own race in the future. That is, it is precisely the profit-maximizing behaviour which leads to equilibrium discrimination.

According to statistical discrimination theories, employers believe that observable physical attributes of the workers are correlated with unobservable but payoff-relevant characteristics. For an overview of statistical discrimination theories, see Fang and Moro (2010). Phelps (1972) explains differences in wages of black and white workers by assuming that the unobservable productivity of a worker is correlated with his colour. Employers use the colour as a signal about the productivity of a worker.

Arrow (1973) shows that discrimination can be a result of self-fulfilling expectations even if all agents are identical ex ante. Workers in his model can decide how much to invest in human capital. These decisions are not observable. Employers expect black workers to invest less than white workers and, hence, they offer lower wages to black workers. Anticipating this, black workers rationally invest less in human capital than white workers. As a result, workers of different colours are different ex-post. Coate and Loury (1993) places Arrow's arguments in an equilibrium model but treats wages exogenously, like our model does. This assumption is relaxed in Moro and Norman (2004). Rosén (1997) offers another explanation for self-fullfilling statistical discrimination. In this model, workers privately observe their idiosyncratic productivity prior to applying for a job. If black workers apply for jobs with lower productivities, firms rationally expect white applicants to be more productive. Therefore, firms prefer to hire white workers which results in a lower value for unemployed black workers. As a consequence, black workers rationally apply for jobs even if they are less productive. In Mailath, Samuelson, and Shaked (2000), employers perfectly observe the worker's productivities. However, the employers may decide not to search for workers among black workers in anticipation of low skill-investment.

The workers in our model are identical both ex-ante and ex-post. A notable difference between statistical discrimination and our theories is that white and black individuals might mutually discriminate each other in our model. Such a phenomenon is inconsistent with statistical discrimination because the signal value of the colour must be the same for an employer regardless of his

own colour.

Lang, Manove, and Dickens (2005) shows that small taste-based or statistical discrimination can have large effects. They show that even if employers have only lexicographic preferences towards hiring white workers, or white workers only slightly more productive then black ones, white workers might end up with significantly higher wages than black ones.

There are a few models where discrimination arises without any payoff-relevant differences between agents of different colours. Eeckhout (2006) considers a dynamic marriage market where individuals are randomly matched. Once a marriage is formed, the parties repeatedly play the Prisoner's Dilemma. If either party defects, both individuals return to the market and are rematched. In order to induce some cooperation, the equilibrium play must involve defection with positive probability at the beginning of a marriage. Otherwise, agents would defect and search for a new partner. The author shows that any colour-blind equilibrium is Pareto dominated by strategies where the probability of defection depends on the colour of the partner.

A white agent discriminates against back workers in our model because of the threat of not being hired by other white agents in the future. The punishment by peers against acting differently from the peers is a well-known phenomenon in sociology as well as in economics, see Austen-Smith and Jr. (2005) and the references therein.

In the model of Mailath and Postlewaite (2006), there is a population of men and women who are matched and produce offspring in each period. Agents differ in their non-storable endowments and care about the consumption of their descendents. In addition, some agents have a physical attribute, such as blue eye-colour. This attribute is inherited by the offspring. There exist equilibria where the attribute has a value, that is, agents with the attribute are better off than agents without it. This is because high-endowment agents without the attribute prefer to match with low-endowment agents with attribute rather than with high-endowment agents with no attribute. The reason is that agents are risk-averse and hence, high-endowment agents are willing to forego present consumption in order to increase the expected consumption of their offspring by equipping them with the attribute. In other words, the biological attribute is used to transfer wealth to future generations.[1] Since agents in our paper are risk-neutral they have no incentive to transfer wealth across periods. However, the social colour in our model is payoff-irrelevant but has a value in equilibria, much like the biological attribute in Mailath and Postlewaite (2006).

The social colour in our model plays a role which is similar to the *labels* in Kandori (1992). Kandori considers a model where members of two communities interact with each others repeatedly. Each member of a community is randomly matched with a member of the other community and plays a game in every period. Players only observe the actions played in their past matches. However, each player carries an observable information, a label, about his past history of actions. A label of an individual is determined by his previous label, his partner's label, and the action he takes. Although the labels are not directly payoff-relevant, players might condition their behaviour on

[1] A similar explanation has been proposed to explain the evolution of peacock tails Ridley (1993)

them. The author proves a Folk Theorem for this setting. Unlike Kandori (1992) and Mailath and Postlewaite (2006), we do not only show that acting on payoff-irrelevant information is a possibility, but we prove that stable equilibria *necessarily* involve discrimination in some environment.

2 The Model

There is a unit mass of population of agents. Each agent lives forever and is risk-neutral. Time is continuous and the common discount rate is r.

Agents randomly receive opportunities to participate in production. These opportunities arrive independently across agents and time according to a Poison distribution with arrival rate δ. Agents with opportunities are matched with one another instantaneously. In each match, one of the agents becomes an *employer* and the other one becomes a *worker* with equal probabilities.[2] The two agents observe a match specific shock, s, which is exponentially distributed, that is, $G(s) = 1 - e^{-\lambda s}$. Then, the employer decides whether or not to employ the worker. If the employer employs the worker, he receives a payoff of s and the worker receives a constant wage, $M\,(>0)$.[3] Otherwise, both agents receive a payoff of zero. Each agent maximizes the discounted present value of monetary payoffs.

Each agent has a two-dimensional type. The first coordinate is the physical colour of the agent and the second one is the *social colour*. The physical colour is either black (b) or white (w) and never changes. A fraction of μ_w of the population is white and a fraction of $\mu_b\,(=1-\mu_w)$ is black. The social colour is also either black or white and evolves as follows. The social colour of a worker does not change.[4] If an employer employs a worker with type (c_1, c_2) then the social colour of the employer remains the same with probability $1-\gamma$, changes to c_1 with probability $\gamma\alpha$ and becomes c_2 with probability $\gamma\,(1-\alpha)$. If the employer decides not to employ, his social colour remains the same with probability $(1-\gamma)$ and becomes his physical colour with probability γ.

Prior to making a decision, an employer observes the type of the worker but nothing else. Note that the social colour of an agent carries information about his past employees. An agent's social colour is more likely to be white if he hired white workers, or workers with white social colour in the past.

The types of the agents are payoff irrelevant in the following sense. The payoff of an agent only depends on the histories of shock realizations and employment decisions but not on his type and the agents' types with whom he interacts. If there were no types, there was a unique equilibrium in this model where employers always employ the workers. In fact, this is true even if agents have physical colours but no social colours. This is because an employer receives a positive payoff if he

[2] Following the convention of the literature on racial discrimination, we adopt the employer-employee terminology. However, we interpret a parnership as any mutually beneficial social or economic interaction.

[3] That is, the total surplus generated in a relationship is $s+M$.

[4] Recall that workers do not make decisions. Any change in the social colour of a worker would be just noise from his point of view. We avoid dealing with this randomness by making this assumption.

employs the worker and in the absence of social colour, such a decision cannot affect his future employment.

In this model, only the employers make decisions. A strategy of an employer is a mapping from his past history, his type, and the type of the worker into an employment decision. In what follows, we restrict attention to *steady state equilibria*. That is, we characterize equilibria where the agents' strategies depend neither on time, nor on history.

3 Best Responses

In this section, we reduce the problem of finding an equilibrium to a two-dimensional one. If an employer with a given type is better off employing a worker given a certain realization of the shock then he would be strictly better off employing the same worker if the realization of the shock was higher. Therefore, the employment decisions can always be characterized by cutoffs. These cutoffs can depend on the types of both the employer and the worker, so there might be sixteen of them. We will show, however, that the employer's social colour does not affect these cutoffs. So, four cutoffs characterize the strategy of a white employer and another four cutoffs define the strategy of a black employer. We will establish a relationship between these cutoffs and the value functions of the agents. This relationship is then used to show that the various cutoffs of a black (white) employer are linearly dependent on one another and the coefficients are determined by the parameters of our model. This implies that the value of one of the cutoffs determines the other three cutoffs. Therefore, each equilibrium is identified by two cutoffs: one for a black employer and one for a white one.

In the rest of this section, we will characterize the equilibrium values in terms of the two cutoffs and express the best-response cutoffs of black and white agents as a function of the cutoffs used by black and white employers. Finally, we derive an explicit formula for these best-response functions and investigate their analytical properties.

3.1 Optimal Cutoffs

We fix a population strategy and distribution of types. We derive the best-response cutoffs of each agent at a certain moment.[5] To this end, let V_{c_1,c_2} denote the value function of an agent with type (c_1, c_2) $\left(\in \{b, w\}^2\right)$ at this moment when he does not have a production opportunity. That is, V_{c_1,c_2} is the maximum discounted present value of the payoffs what a type-(c_1, c_2) agent can achieve given the strategy and type-distribution of the others.[6]

Next we compute the optimal cutoff of a white employer with social colour c who faces a worker with type (b, w) at this moment. Suppose that the value of the shock is s. If he employs the worker,

[5] We emphasize that the population strategies do not have to be equilibrium strategies.

[6] Since two agents with the same type face the same environment, their values are the same.

he receives an instantaneous payoff of s. His social colour remains c with probability $(1-\gamma)$ and changes to b or w with probabilities $\gamma\alpha$ and $\gamma(1-\alpha)$, respectively. Hence, if the worker is hired, the discounted present value of the employer's payoffs is

$$s + (1-\gamma) V_{w,c} + \gamma\alpha V_{w,b} + \gamma(1-\alpha) V_{w,w}. \tag{1}$$

If he does not employ the worker, his value was

$$(1-\gamma) V_{w,c} + \gamma V_{w,w}. \tag{2}$$

The employer is better off hiring the worker whenever (1) is larger than (2). The cutoff, above which the worker is employed, is the shock realization, s, which makes (1) and (2) equal. That is, the best-response cutoff is $\gamma\alpha(V_{w,w} - V_{w,b})$. Since the shock is always positive, having a negative cutoff is equivalent to having a zero cutoff. Therefore, one can restrict attention to weakly positive cutoffs, in which case, the best-response cutoff is uniquely defined by $\max\{0, \gamma\alpha(V_{w,w} - V_{w,b})\}$.

Notice that this cutoff does not depend on the social colour of the employer, c. In both (1) and (2), the only term which depends on c is $(1-\gamma) V_{w,c}$, and therefore, c cancels out in the computation of the cutoff. In fact, the cutoff of an agent never depends on his social colour in equilibrium. This is because the social colour of an employer only affects his payoff in the event when his new social colour remains his old one, and this event is independent of his decision. Therefore, the equilibrium cutoff of an agent can only depend on his physical colour but not on his social colour.

Let $x^c_{c_1,c_2}$ denote the cutoff value of an employer with physical colour c if the type of the worker is (c_1, c_2). We denote the colour which is not c by $-c$ for $c \in \{w, b\}$. Above, we have shown that $x^w_{b,w} = \max\{0, \gamma\alpha(V_{w,w} - V_{w,b})\}$. The other cutoffs can be computed similarly and they are summarized by the following

Lemma 1 *The following equations establish the relationship between best-response cutoffs and the value functions:*

$$\begin{aligned}
x^c_{-c,-c} &= \max\{0, \gamma(V_{c,c} - V_{c,-c})\}, \\
x^c_{c,-c} &= \max\{0, \gamma(1-\alpha)(V_{c,c} - V_{c,-c})\}, \\
x^c_{-c,c} &= \max\{0, \gamma\alpha(V_{c,c} - V_{c,-c})\}, \\
x^c_{c,c} &= 0.
\end{aligned}$$

If an employer with physical colour c is considering hiring a worker, he is concerned about his new social colour. Having a social colour c instead of $-c$ provides the agent with an additional value of $V_{c,c} - V_{c,-c}$. This difference can be interpreted as a *bias* of the agent towards his own physical colour.[7] The lemma says that the best-response cutoffs are proportional to this bias up to the requirement that the cutoffs are non-negative. The coefficients of the bias corresponding to

[7] This bias may well be negative, that is, an agent is better off if his physical colour does not coincide with his social colour.

various cutoffs are determined by the probabilities of the social colour becoming c and $-c$, which in turn, depend on the type of the worker.

Let $x^c = x^c_{-c,-c}$ and notice that

$$x^c_{c,-c} = (1-\alpha)\, x^c,\ x^c_{-c,c} = \alpha x^c, \text{ and } x^c_{c,c} = 0. \tag{3}$$

Since the value functions of two agents with the same type are identical, this lemma implies that any stationary equilibrium is symmetric. That is, employers with the same physical colour use the same strategies. Also notice that, by (3), an equilibrium strategy of a colour-c employer is identified by x^c. In what follows, we refer to the cutoff x^c as a strategy (or cutoff) while keeping in mind that the cutoffs against different types of workers are defined by (3).

3.2 The Best-Response Curves

The goal of this section is to explicitly characterize the best responses of black and white agents as functions of the cutoffs of others. We denote the best response cutoff of an agent with colour c by $b^c\left(x^c, x^{-c}\right)$ if each employer with physical colour c $(-c)$ uses cutoff x^c (x^{-c}).

The main result of this section is stated in the following

Lemma 2 *The best response curve of an agent with colour c is defined by the following equation:*

$$b^c\left(x^c, x^{-c}\right) = \frac{M\delta\gamma}{2r} \max\left\{0, K\left[\mu_c G\left((1-\alpha)\, x^c\right) + \mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)\right]\right\}. \tag{4}$$

The rest of this section is devoted for the proof of this lemma. In fact, we do not only characterize best responses against constant strategies where each colour-c agent uses the same cutoff but against any stationary distribution of strategies as long as these strategies satisfy the equations in (3). This turns out to be useful when we examine the stability properties of the equilibria later. Let X^c $(c \in \{b, w\})$ denote the random variable corresponding to the distribution of cutoffs of colour-c agents in the population. We shall compute $V_{c,c} - V_{c,-c}$ for $c \in \{b, w\}$ conditional on $\left(X^b, X^w\right)$.[8] These objects then identify the best-response cutoffs by Lemma 1.

Let $\Pi^l_{c_1,c_2}$ and $\Pi^e_{c_1,c_2}$ denote the value functions of a worker and employer with type $(c_1, c_2) \in \{b, w\}^2$, respectively. The heuristic equation describing the relationship between V_{c_1,c_2}, $\Pi^l_{c_1,c_2}$ and $\Pi^e_{c_1,c_2}$ is:

$$V_{c_1,c_2} = (1-\delta dt)\,(1-r dt)\, V_{c_1,c_2} + \delta dt \left(\frac{1}{2}\Pi^s_{c_1,c_2} + \frac{1}{2}\Pi^e_{c_1,c_2}\right).$$

To see this, notice that an agent does not receive an opportunity with probability $1-\delta dt$ in the next dt time, and hence his value remains V_{c_1,c_2}. This is discounted at the rate r. With the remaining probability, the agent receives an opportunity and becomes an employer or a worker with equal probabilities. Dividing through by dt and taking the limit as dt goes to zero:

$$V_{c_1,c_2} = \frac{\delta}{\delta + r}\left(\frac{1}{2}\Pi^l_{c_1,c_2} + \frac{1}{2}\Pi^e_{c_1,c_2}\right). \tag{5}$$

[8] Of course, the values of the agents depend on $\left(X^b, X^w\right)$. This dependence is supressed from the notation V_{c_1,c_2} for simplicity.

A worker with type (c,c) is employed whenever he is matched with an employer with physical colour c, which happens with probability μ_c. He also gets employed whenever he is matched with an employer with physical colour $-c$ whose cutoff is x^{-c} and $s \geq x^{-c}$. This happens with probability $\mu_{-c}\left(1 - EG\left(X^{-c}\right)\right)$, where the expectation is taken according to the distribution of the cutoff X^{-c}. Finally, the worker's value changes to $V_{c,c}$ and gets M whenever he is employed, therefore,

$$\Pi^l_{c,c} = M\left(\mu_c + \mu_{-c}\left(1 - EG\left(X^{-c}\right)\right)\right) + V_{c,c}. \tag{6}$$

Similarly,

$$\Pi^l_{c,-c} = M\left(\mu_c\left(1 - EG\left(\left(1-\alpha\right)X^c\right)\right) + \mu_{-c}\left(1 - EG\left(\alpha X^{-c}\right)\right)\right) + V_{c,-c}. \tag{7}$$

Using (5), (6), and (7) we can express $V_{c,c} - V_{c,-c}$ as follows:

$$\begin{aligned} V_{c,c} - V_{c,-c} &= \frac{\delta}{\delta + r}\left[\frac{1}{2}\left(\Pi^l_{c,c} - \Pi^l_{c,-c}\right) + \frac{1}{2}\left(\Pi^e_{c,c} - \Pi^e_{c,-c}\right)\right] \\ &= \frac{\delta}{\delta + r}\frac{1}{2}M\left[\mu_c EG\left(\left(1-\alpha\right)X^c\right) + \mu_{-c}\left(EG\left(\alpha X^{-c}\right) - EG\left(X^{-c}\right)\right)\right] \\ &\quad + \frac{\delta}{\delta + r}\left[V_{c,c} - V_{c,-c}\right] \end{aligned}$$

That is,

$$V_{c,c} - V_{c,-c} = \frac{M\delta}{2r}\left[\mu_c EG\left(\left(1-\alpha\right)X^c\right) + \mu_{-c}\left(EG\left(\alpha X^{-c}\right) - EG\left(X^{-c}\right)\right)\right].$$

Recall from Lemma 1 that the best-response cutoff of an employer with physical colour c against a worker with type $(-c,-c)$ is $\gamma\left(V_{c,c} - V_{c,-c}\right)$. Let $K = M\delta\gamma/2r$. Then substituting in from the previous displayed equality:

$$K\left[\mu_c EG\left(\left(1-\alpha\right)X^c\right) + \mu_{-c}\left(EG\left(\alpha X^{-c}\right) - EG\left(X^{-c}\right)\right)\right]. \tag{8}$$

Suppose now that each employer with physical colour c uses x^c, that is, $X^c \equiv x^c$. Then the best response of an agent can be written as

$$\widetilde{b}^c\left(x^c, x^{-c}\right) = K\left[\mu_c G\left(\left(1-\alpha\right)x^c\right) + \mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)\right]. \tag{9}$$

Recall that since the shocks are always positive, one can restrict attention to weakly positive cutoffs, in which case, the best-response correspondence is uniquely identified by

$$b^c\left(x^c, x^{-c}\right) = \max\left\{0, \widetilde{b}^c\left(x^c, x^{-c}\right)\right\},$$

which is just (4). A notable feature of the best response function is that it does not depend on the distribution of social colours.

3.3 Properties of the Best-Response Curves

The next two lemmas describe some properties of the best-response curves.

Lemma 3 *The function b^c satisfies the following properties:*
(i) for each x^{-c}, if $b^c(x^c, x^{-c}) > 0$ then b^c is locally concave in x^c,
(ii) $b^c(0, x^{-c}) = 0$ for all x^{-c},
(iii) for all $\overline{x}^{-c} > 0$, $b^c(x^c, 0) = \lim_{x^{-c}\to\infty} b^c(x^c, x^{-c}) \geq b^c\left(x^c, \overline{x}^{-c}\right)$.

Part (ii) says that a colour-c agent's best-response cutoff is zero if the cutoff of each colour-c agent is also zero. An immediate implication of this observation is that the colour blind strategy constitutes an equilibrium. We state it in the next

Remark 1 *The cutoff profile defined by $x^c = x^{-c} = 0$ is an equilibrium.*

Before we provide a graphical representation of this lemma, we analyse the first derivative of b^c. Note that part (ii) implies that the function $b^c(x^c, 0)$ intersects with the 45 degree line at $x^c = 0$. Whether or not there is another intersection depends on this derivative and has great importance in characterizing the set of equilibria.

Lemma 4 *Let $\lambda_0 = 1/(K(1-\alpha)\mu_c)$. Then,*
(i) if $\lambda > \lambda_0$, then there exists a unique $x^c > 0$ such that $b^c(x^c, 0) = x^c$, and
(ii) if $\lambda \leq \lambda_0$, then $b^c(x^c, 0) < x^c$ for all $x^c > 0$.

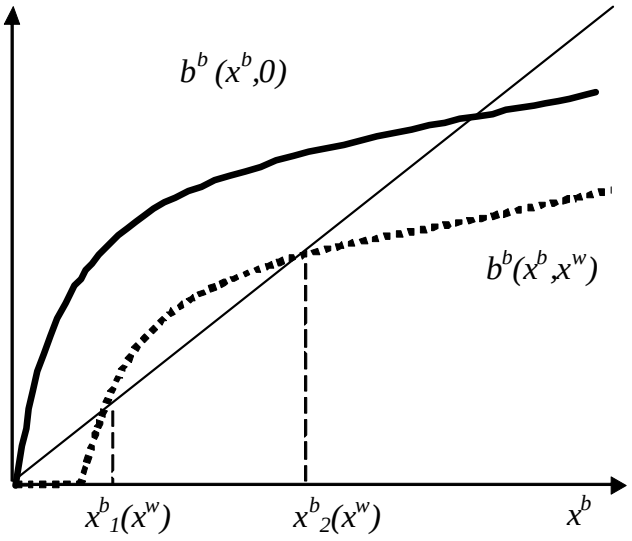

Figure 1: Best Responses

Figure 1 plots $b^b\left(x^b,0\right)$ and $b^b\left(x^b,x^w\right)$ $(x^w>0)$ for the case of a large λ. The function $b^b\left(x^b,0\right)$ is identical to $\widetilde{b}^b\left(x^b,0\right)$ because, in this case, $\widetilde{b}^b$ is weakly positive (see (4) and (9)). For $x^w>0$, $b^b\left(x^b,x^w\right)$ is a downwards shift of $b^b\left(x^b,0\right)$ except it is zero whenever the shifted curve was negative. Since $\widetilde{b}^c$ is concave in x^b, $b^b\left(x^b,x^w\right)$ is locally concave in x^w whenever it is positive (part (i) of Lemma 3). Part (ii) of Lemma 3 says that if the cutoff of each black agent is zero then the best response cutoff of a black agent is also zero. To see this, notice that if $x^b=0$ then black agents are better off having a white social colour than a black one. This is because their social colours have no impact on their employment if the employer is black $\left(x^b=0\right)$ but they are more likely to be employed by white agents if their social colour is white. Therefore, a black employer always employs a type-(w,w) worker, that is, the best-response cutoff is zero.

Part (iii) of Lemma 3 states that the best response cutoff of a black agent is the same when white agents do not discriminate $(x^w=0)$ or they fully discriminate $(x^w=\infty)$. To see this note that a black agent is always employed by white agents if $x^w=0$ and is never employed by them if $x^w=\infty$. That is, the white agents' decisions to hire black workers do not depend on the workers' social colours. Therefore, the black workers' best-response is determined solely by the cutoff x^b in both cases.

Part (iii) also says that b^b decreases if x^w becomes larger than zero. The intuition is as follows. As x^w becomes positive, a black worker benefits from having a white social colour whenever he meets a white employer. Therefore, holding x^b fixed, a black agent has less incentive to discriminate against type-(w,w) workers, that is, b^b goes down.

Note that the curve $b^b\left(x^b,0\right)$ intersects with the 45-degree line twice in Figure 1. The function $b^b\left(.,0\right)$ is strictly concave and zero at zero. In addition, this slope converges to zero as x^b goes to infinity. Therefore, the function $b^b\left(.,0\right)$ intersects with the 45-degree line at a strictly positive value if and only if its slope is larger than one at zero. The slope of $b^b\left(.,0\right)$ is large if and only if $\lambda$is large (Lemma 4).

Next, we argue that whenever the positive intersection of $b^c\left(.,0\right)$ and the 45-dgeree line exists, that is if λ is large, there exists at least one equilibrium which involves discrimination. Suppose for a moment that agents with physical colour $-c$ are non-strategic and their cutoff is zero, and consider our model as a game played by only agents with colour c. Part (i) of Lemma 4 implies that there exists an $x^c\left(>0\right)$ such that $b^c\left(x^c,0\right)=x^c$, that is, the best response of an agent is x^c whenever every other agent with colour c uses cutoff x^c. In other words, the cutoff x^c is an equilibrium in the game where only colour-c agents act strategically. Since this cutoff is positive, agents discriminate against others with physical or social colour $-c$. What if agents of colour $-c$ become strategic? Part (ii) of Lemma 3 implies that zero is a best response of an agent with colour $-c$ as long as $x^{-c}=0$. That is, if the population strategy is described by the cutoff profile is $(x^c,0)$ then each agent best responds. We can conclude the following

Remark 2 *If $\lambda>1/\left(K\left(1-\alpha\right)\mu_c\right)$ then for each $c\in\{b,w\}$, there is a unique $x^c>0$ such that the cutoff profile $(0,x^c)$ is an equilibrium.*

4 Equilibrium Characterization

This section describe some qualitative features of the equilibria for the case of large λ.

The definition of equilibrium can be restated in terms of the best-response curves as follows. The cutoff profile (x_*^c, x_*^{-c}) is an equilibrium if and only if

$$\left(x_*^c, x_*^{-c}\right) = \left(b^c\left(x_*^c, x_*^{-c}\right), b^{-c}\left(x_*^{-c}, x_*^c\right)\right). \tag{10}$$

In particular, $x_*^c = b^c\left(x_*^c, x_*^{-c}\right)$ for $c \in \{b, w\}$. This means that the function $b^c\left(., x_*^{-c}\right)$ intersects with the 45-degree line at x_*^c. In what follows, we analyse the intersections of the curve $b^c\left(., x^{-c}\right)$ and the 45-degree line for each x^{-c}. As we pointed out (see part (ii) of Lemma 3), these curves intersect at zero. Next, we investigate the strictly positive intersections.

Recall that the curve $b^c\left(., x^{-c}\right)$ is essentially a downward shift of $b^c\left(., 0\right)$ (see Figure 1). The size of this shift determines the number of positive intersections. We will show that this size is a non-monotonic function of x^{-c}. If x^{-c} is small, an increase in x^{-c} shifts down the curve $b^c\left(., x^{-c}\right)$ even more. Above a certain value of x^{-c}, however, a further increase in x^{-c} shifts the curve $b^c\left(., x^{-c}\right)$ upwards. In fact, as x^{-c} goes to infinity, $b^c\left(., x^{-c}\right)$ converges back to $b^c\left(., 0\right)$ (see part (iv) of Lemma 3). Recall that if λ is large, the first derivative of $b^c\left(., 0\right)$ is larger than one (see Lemma 4). Hence, $b^c\left(., x^{-c}\right)$ and the 45-degree line have two intersections if the downward shift is small and zero if the shift is large.[9] In the latter case, $b^c\left(., x^{-c}\right)$ is pushed below the 45-degree line. The former case, where there are two intersections, is depicted on Figure 1 for $c = b$. (The intersections are denoted by $x_1^b\left(x^w\right)$ and $x_2^b\left(x^w\right)$.) We will prove that there are two cases that can arise: Case 1: the curve $b^c\left(., x^{-c}\right)$ intersects with the 45-degree line even when it is shifted down the most. Case 2: there is an interval such that the curve $b^c\left(., x^{-c}\right)$ is pushed below the 45-degree line if x^{-c} lies in this interval. If x^{-c} is outside of this interval, there are two positive intersections.

Figure 2 illustrates these two cases. In this figure, $x_1^c\left(x^{-c}\right)$ and $x_2^c\left(x^{-c}\right)$ denote the smaller and larger positive intersections, respectively. In Case 1, these intersections always exist. In Case 2, the curve $b^c\left(., x^{-c}\right)$ is below the 45-degree line if $x^{-c} \in \left(\underline{x}^{-c}, \overline{x}^c\right)$ and $x_1^c\left(x^{-c}\right)$ and $x_2^c\left(x^{-c}\right)$ are not defined on this interval. In both cases, the curve x_1^c increases first, then decreases. This is because as the downward shift gets larger (smaller) the place of the first positive intersection increases (decreases). Similarly, the curve x_2^c decreases first, then increases because the place of the second positive intersection decreases (increases) as the shift gets larger (smaller). In the panel corresponding to Case 2, the values of x_1^c and x_2^c are equal at $\underline{x}^{-c}$ and $\overline{x}^c$. The reason is that both $\underline{x}^{-c}$ and $\overline{x}^c$ induce the same shift, that is, $b^c\left(., \underline{x}^{-c}\right) = b^c\left(., \overline{x}^c\right)$. In addition, the shifted best-response curve is exactly tangent to the 45-degree line, hence, the two intersections collapse into one.

We state these results in the next lemma and prove them in the Appendix.

[9] There is a non-generic third case where the curve $b^c\left(., x^{-c}\right)$ is tangent to the 45-degree line when it is shifted down the most.

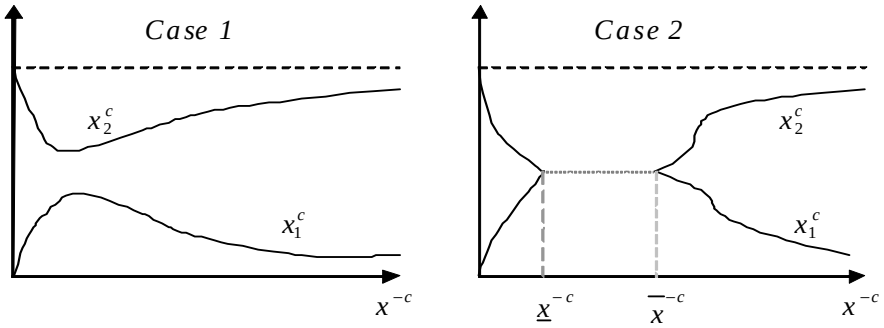

Figure 2: Positive Intersections

Lemma 5 *If $\lambda \geq 1/\left(K\left(1-\alpha\right)\mu_c\right)$ then either*

(i) for all $x^{-c} > 0$ there exist $x_1^c\left(x^{-c}\right)$, $x_2^c\left(x^{-c}\right)$ such that $x_i^c\left(x^{-c}\right) = b^c\left(x^c\left(x^{-c}\right), x^{-c}\right)$ and $0 < x_1\left(x^{-c}\right) < x_2\left(x^{-c}\right)$, or

(ii) there exist $\underline{x}^{-c}$,$\overline{x}^{-c} \in \mathbb{R}_{++}$, $\underline{x}^{-c} \leq \overline{x}^{-c}$, such that for all $x^{-c} \in \left(\underline{x}^{-c}, \overline{x}^{-c}\right)$: $b^c\left(x^c, x^{-c}\right) < x^c$, and for all $x^{-c} \in \mathbb{R}_{++} \setminus \left[\underline{x}^{-c}, \overline{x}^{-c}\right]$ there exist $x_1^c\left(x^{-c}\right)$, $x_2^c\left(x^{-c}\right)$ such that $x_i^c\left(x^{-c}\right) = b^c\left(x_i^c\left(x^{-c}\right), x^{-c}\right)$, $0 < x_1\left(x^{-c}\right) < x_2\left(x^{-c}\right)$, and

$$\lim_{x^{-c}\to\underline{x}^{-c}} x_1^c\left(x^{-c}\right) = \lim_{x^{-c}\to\overline{x}^{-c}} x_1^c\left(x^{-c}\right) = \lim_{x^{-c}\to\overline{x}^{-c}} x_2^c\left(x^{-c}\right) \lim_{x^{-c}\to\underline{x}^{-c}} x_2^c\left(x^{-c}\right).$$

In addition, $x_1^c\left(x^{-c}\right)$ is increasing first, then is decreasing, and $x_2^c\left(x^{-c}\right)$ is decreasing first, then is increasing. Finally, $\lim_{x^{-c}\to 0} x_1^c\left(x^{-c}\right) = 0$.

The curves x_1^c and x_2^c are only defined for strictly positive values of x^{-c}. It turns out to be useful to define $x_i^c\left(0\right) = \lim_{x^{-c}\to 0} x_i^c\left(x^{-c}\right)$. Note that $x_1^c\left(0\right) = 0$ and $x_2\left(0\right)$ corresponds to the positive intersection of $b^c\left(.,0\right)$ and the 45-degree line. In addition, since $b^c\left(.,x^{-c}\right)$ intersects with the 45-degree line at zero for all x^{-c} (see part (ii) of Lemma 3), the curve $x_0^c\left(x^{-c}\right) \equiv 0$ also defines an intersection.

Now, we can define equilibria in terms of the intersections of the curves $\left\{x_i^b\right\}_{i=0}^2$ and $\left\{x_i^w\right\}_{i=0}^2$. Formally, (x_*^c, x_*^{-c}) is an equilibrium cutoff profiles if and only if there exist $i, j \in \{0,1,2\}$ such that

$$x_*^c = x_i^c\left(x_*^{-c}\right) \text{ and } x_*^{-c} = x_j^{-c}\left(x_*^c\right). \tag{11}$$

Therefore, in order to find equilibria geometrically, we need to add the curves $\left\{x_i^{-c}\right\}_{i=0}^2$ on Figure 2 and find all the intersections. We did just this on Figure 3 corresponding to Case 1 in Figure 2.

Notice that by (9) the best-response cutoff of an agent with colour c is largest if $x^c = \infty$ and $x^{-c} = 0$. In this case, the best-response cutoff is $K\mu_c$. This implies that the equilibrium cutoff of

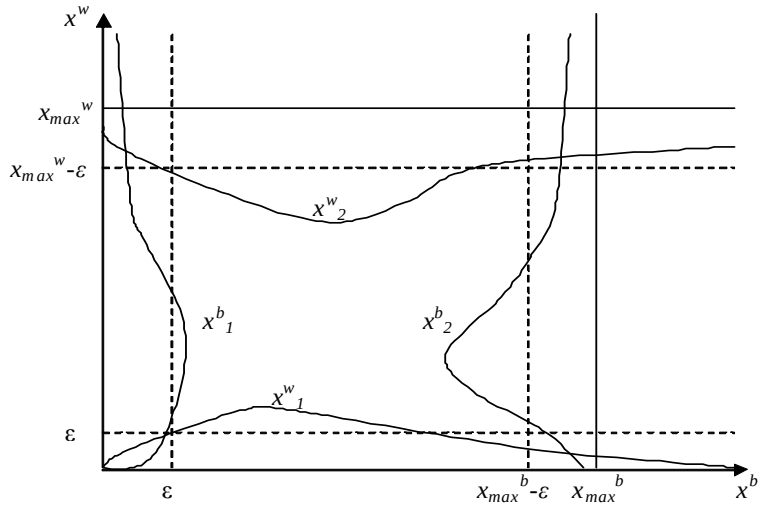

Figure 3: Equilibria

an agent with colour c can never exceed $K\mu_c$. Let $x^c_{\max} = K\mu_c$. We are ready to state the main result of this section.

Proposition 1 *For all K, μ_c, α and $\varepsilon\,(>0)$ there exists a λ_0 such that if $\lambda \geq \lambda_0$, then if x^c_* is an equilibrium cutoff then either:*

(i) $x^c_* = 0$, or

(ii) $x^c_* \in (0, \varepsilon)$, or

(iii) $x^c_* \in (x^c_{\max} - \varepsilon, x^c_{\max})$.

This proposition states that if λ is large enough then, in every equilibrium, an agent either does not discriminate at all ($x^c_* = 0$), or weakly discriminates ($x^c_* < \varepsilon$), or strongly discriminates ($x^c_{\max} - \varepsilon < x^c_*$). Proposition 1 is illustrated on Figure 3. Intuitively, strong discrimination of agents with colour c corresponds to the curve x^c_2, weak discrimination corresponds to x^c_1, and x^c_0 implies no discrimination.

The statement of the proposition allows any combination of these possibilities in equilibrium. In Remark 2, we have shown that the intersection of x^c_0 and x^{-c}_2 exists and is unique ($c \in \{b, w\}$). The unique intersection of x^c_0 and x^{-c}_0, $(0,0)$, corresponds to the colour-blind equilibrium. Since $x^c_1(0) = 0$, the intersection of x^c_1 and x^{-c}_0 is $(0,0)$, that is, there is no equilibrium where one colour weakly discriminates and the other does not discriminate at all. The proposition neither implies the existence, nor the uniqueness of any of the other types of equilibria. The next section introduces a stability concept and we shall fully characterize those equilibria which are stable.

5 Stability

Next, we introduce a fairly standard notion of stability[10]. It is based on the requirement that a simple myopic best-response dynamics converges to the equilibrium if the strategies of the agents are slightly perturbed around the equilibrium. We model the myopic best-response dynamics by assuming that initially agents best-responde to some stationary population strategy which might be different from the actual one. Then each agent stochastically receivesan opportunity to update his strategy. Whenever an agent has this opportunity, he myopically adjusts his strategy to the current environment. That is, he best-responds to the current population strategy as if it was never to change.

Formally, suppose that initially the distribution of cutoffs of black and white agents are described by a pair of random variables $\left(X^b, X^w\right)$ and the strategy of each agent satisfies the statement of Lemma 1.[11] Agents receive opportunities to update their strategies according to a Poison process with an arrival rate normalized to be one.[12] If an agent has this opportunity at time t, he best-responds to the cutoff distribution at t as if it was constant over time. Let $x_t^c\left(X^c, X^{-c}\right)$ denote the best-response cutoffs of an agent with colour c at time t if the initial distribution of cutoffs was (X^c, X^{-c}).

Definition 1 *The equilibrium cutoff vector* $\left(x_*^b, x_*^w\right)$ *is said to be* stable *if there exists an* $\varepsilon > 0$*, such that if* $|X^c - x_*^c| < \varepsilon$ *almost surely for* $c \in \{b, w\}$ *then* $\lim_{t\to\infty} x_t^c\left(X^c, X^{-c}\right) = x_*^c$ *for* $c \in \{b, w\}$*.*

In what follows we describe the equation governing the best-response dynamics. Fix $\left(X^b, X^w\right)$ and let $\left(X_t^b, X_t^w\right)$ denote the distribution of population cutoffs at time t. We shall denote the best response of an agent with colour c by x_t^c suppressing its argument $\left(X^b, X^w\right)$. By (8), the best-response of an agent with colour c at time t is

$$x_t^c = K\left[\mu_c EG\left(\left(1-\alpha\right)X_t^c\right) + \mu_{-c}\left(EG\left(\alpha X_t^{-c}\right) - EG\left(X_t^{-c}\right)\right)\right].$$

Next, we approximate x_{t+dt}^c by assuming that between t and $t+dt$ each agent changes his strategy to the time t best-response cutoffs. There is a measure of dt agents who receive an opportunity to change their strategies between t and $t+dt$ and they all switch to x_t^c. Therefore,

$$\begin{aligned}
x_{t+dt}^c &= K\left[\mu_c EG\left(\left(1-\alpha\right)X_{t+dt}^c\right) + \mu_{-c}\left(EG\left(\alpha X_{t+dt}^{-c}\right) - EG\left(X_{t+dt}^{-c}\right)\right)\right] \\
&= \left(1-dt\right)K\left[\mu_c EG\left(\left(1-\alpha\right)X_t^c\right) + \mu_{-c}\left(EG\left(\alpha X_t^{-c}\right) - EG\left(X_t^{-c}\right)\right)\right] \\
&\quad + dtK\left[\mu_c EG\left(\left(1-\alpha\right)x_t^c\right) + \mu_{-c}\left(EG\left(\alpha x_t^{-c}\right) - EG\left(x_t^{-c}\right)\right)\right] \\
&= \left(1-dt\right)x_t + dt\widetilde{b}^c\left(x_t^c, x_t^{-c}\right),
\end{aligned}$$

[10] See, for example, Chapter 3 of Fudenberg and Levine.

[11] This latter assumption is satisfied if each agent best responds to *some* population strategy.

[12] This normalization is without the loss of generality because this arrival rate only affects the speed of convergence but not the limits.

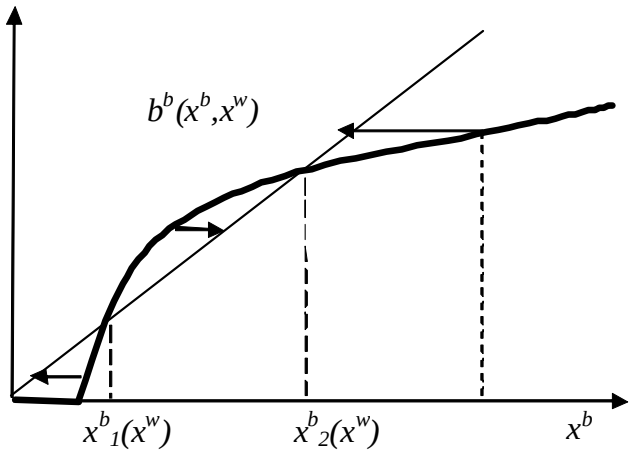

Figure 4: Best-response Dynamics

where the first equality follows because the strategy of $1-dt$ measure of the population is described by $\left(X_t^c, X_t^{-c}\right)$, and the rest uses $\left(x_t^c, x_t^{-c}\right)$. The second equality follows from (4). Taking dt to zero leads to the following differential equation describing the evolution of x_t^c:

$$\frac{dx_t^c}{dt} = \widetilde{b}^c\left(x_t^c, x_t^{-c}\right) - x_t^c.$$

As we mentioned before, we can restrict attention to non-negative cutoffs. Note that if $x_t^c = 0$ then $\widetilde{b}^c\left(x_t^c, x_t^{-c}\right) = b\left(x_t^c, x_t^{-c}\right) = 0$ by (4) and part (ii) of Lemma 3. Hence, the previous displayed equation implies that $dx_t^c/dt = 0$ whenever $x_t^c = 0$. Therefore,

$$\frac{dx_t^c}{dt} = \begin{cases} 0 & \text{if } x_t^c = 0 \\ \widetilde{b}^c\left(x_t^c, x_t^{-c}\right) - x_t^c & \text{if } x_t^c > 0 \end{cases}. \tag{12}$$

Figure 4 helps to understand the best-response dynamics derived from (12). Consider $x^b > x_2\left(x^w\right)$. At this point, the best response curve is below the 45 degree line, that is $b^b\left(x^b, x^w\right) < x^b$. In general, if $\left(x^b, x^w\right)$ is to the right from the x_2^b curve, the best response of a black agent is smaller than x^b. Equation (12) implies that the best response of a black agent decreases on this region. This is represented by a horizontal arrows pointing to the left. A similar argument shows, that if $x_1^b\left(x^w\right) < x^b < x_2^b\left(x^w\right)$, then $b^b\left(x^b, x^w\right) > x^b$. Hence, by (12), the best response of a black agent increases. This is represented by horizontal a arrow pointing to the right between the points $x_1^b\left(x^w\right)$ and $x_2^b\left(x^w\right)$. Finally, if $x^b < x_1^b\left(x^w\right)$, then $b^b\left(x^b, x^w\right) < x^b$ which is represented by a horizontal arrow pointing to the left.

We are ready to state the main theorem of the paper.

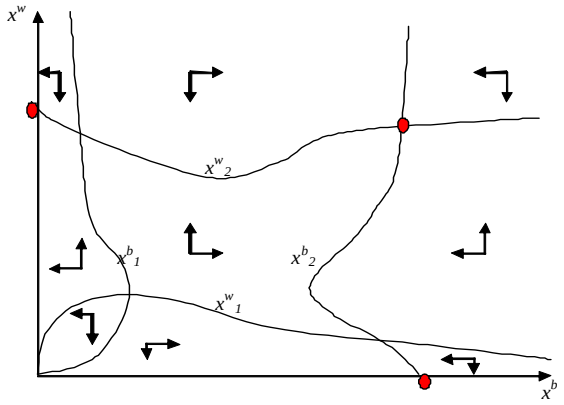

Figure 5: Stable Equilibria

Theorem 1 *For all K, μ_c, α and $\varepsilon\,(>0)$ there exists a λ_0 such that if $\lambda \geq \lambda_0$, then there are exactly three stable equilibria $(x^w_*, 0)$, $\left(0, x^b_*\right)$, and $\left(x'^w_*, x'^b_*\right)$ such that $x^c_*, x'^c_* \in (x^c_{\max} - \varepsilon, x^c_{\max})$ for $c \in \{b, w\}$.*

If λ is large enough there are the following three equilibria in our model. First, the white population discriminates strongly against the black one, while the blacks do not discriminate at all. Second, the blacks strongly discriminate against the whites, while the whites do not discriminate at all. And finally, both population strongly discriminate against the other.

Equation (12) implies that the change in best responses at time t depends only on the time-t best responses, $\left(x^b_t, x^w_t\right)$, but not directly on the distribution of strategies, $\left(X^b_t, X^w_t\right)$. In particular, the initial distribution of strategies affects the best-response dynamics only through the initial best-response profile $\left(x^b_0, x^w_0\right)$. Therefore, using (12), we can represent the best-response dynamics by constructing a Phase Diagram, plotted on Figure 4. For each $\left(x^b, x^w\right)$ cutoff vector, the horizontal and vertical arrows on this figure indicate the directions of the best responses of black and white agents, respectively.

To explain how the arrows are drawn, recall from Figure 3, that if $x^b > x_2\left(x^w\right)$ then $b^b\left(x^b, x^w\right) < x^b$ which is presented by an arrow pointing to the left. In general, if $\left(x^b, x^w\right)$ is to the right from the x^b_2 curve, the best response of a black agent is smaller than x^b. This is represented in Figure 4 by horizontal arrows pointing to the left on the area which is right of the curve x^b_2. Similarly, Figure 3 shows that if $x^b_1\left(x^w\right) < x^b < x^b_2\left(x^w\right)$, then the best response of a black agent increases. This is why the horizontal arrows are pointing to the right between the curves x^b_1 and x^b_2 on Figure 4. Finally, if $x^b < x^b_1\left(x^w\right)$, then $b^b\left(x^b, x^w\right) < x^b$ which is represented by horizontal arrows pointing

to the left on the area which is left of x_1^b. The vertical arrows are constructed in a similar manner representing the best-response dynamics of the white population.

Figure 4 can be used to understand the stability properties of various equilibria. Consider, for example, the colour-blind equilibrium, $(0,0)$. There are points below the curve x_1^w and right to the curve x_1^b arbitrarily close to $(0,0)$. At these points, x_t^b increases and x_t^w decreases, and the vector $\left(x_t^b, x_t^w\right)$ converges to the intersection of x_2^b and x_0^w. Hence, the colour blind equilibrium is unstable. Similarly, it is easy to see that the intersections of the curves x_1^c and x_1^{-c} and the curves x_1^c and x_2^{-c} are unstable for $c \in \{b, w\}$. On the other hand, from any point close to any of the equilibria described in the statement of Theorem 1, the arrows point towards the equilibrium, and hence, these equilibria are stable.

6 Discussion

In this section, we first derive some comparative static results. Then we discuss some of the assumptions and extentions of our model. Finally, compare our empirical implications with those of statistical and taste-based theories of discrimination.

6.1 Comparative Statics

In what follows we focus on the case of a large λ, that is, where the statement of Theorem 1 is valid. Recall that there are exactly three stable equilibria. Next, we investigate how the cutoffs in these three equilibria change if the parameters of our model change. We emphasize that these comparative static results are only valid as long as the change in the parameters are small enough, so that Theorem 1 holds. The following table summarizes the comparative static results. This table indicates what happens to the cutoffs if a certain parameter increases. It turns out that direction of the change in cutoffs is the same in all three equilibria.

	x_*^w	x_*^b
discount rate (r)	down	down
wage (M)	up	up
shock distribution (λ)	up	up
measure of whites (μ_w)	up	down
matching frequency (δ)	up	up
persistence (γ)	up	up

The proof of these results is straightforward, hence, it is omitted. Below, we provide some intuition behind these observations.

An agent cares about his social colour because it influences his future employment. Therefore, the larger is the part of an agent's payoff due to future wages the more likely it is that employer's condition their decisions on the types of the workers. For example, if the discount rate goes up,

agents care more about their current payoffs relative to their future payoffs. Hence, they become more eager to employ workers no matter what their types are, and as a result, their cutoffs go down. Similarly, if M goes up being an employed worker becomes more important, and the cutoffs go up. When λ goes up, the expected payoff of an employer decreases. In other words, having the option of being a worker becomes more important relative to being an employer. This is why an increase in λ has the same effect as a decrease in M.

When μ_w increases, agents receive a larger fraction of their payoffs from interacting with white agents. This makes it more expensive to discriminate against whites, and cheaper to discriminate against blacks. As a result, x_*^w goes up while x_*^b goes down.

An employer's social colour does remains the same with probability $1-\gamma$ independently of his decision. The larger is γ the more likely it is that the worker's type has an impact on the employer's future payoff. Hence, it becomes easier to discriminate.

6.2 Assumptions

Our goal in this paper was to present a simple model which demonstrate that discrimination can arise purely because agents carry information about their past actions. Some of the assumptions are made in order to be able to provide a a graphical representation of equilibria and stability. We do not claim that equilibrium discrimination is robust to all the features of our model. In what follows, we discuss some of our assumptions and extentions of the model.

Distribution of shocks.– We have assumed that the distribution of the match-specific shock which determines the surplus of a partnership is exponential. To what extent does our main result depend on this assumption? For general distributions, we have no hope for a full characterization of equilibria such as in Theorem 1. However, whether or not the colour blind equilibrium is stable depends only on the slope of the best-response functions at $(0,0)$. If this slope is less than one, the colour blind equilibrium is unique and stable. Otherwise, each stable equilibrium involves discrimination. We formally state this result in the following

Theorem 2 *Suppose that s is distributed on $\mathbb{R}_+$ according to the CDF G. If G is concave on $\mathbb{R}_+$ then either*

(i) $(0,0)$ *is the unique equilibrium and is stable, or*

(ii) $(0,0)$ *is not stable, and there exists a stable equilibrium.*

Notice that the total surplus of a partnership, $s+M$, is always positive. This assumption makes the socially optimal employment decisions very easy to characterize. Efficiency requires employers to hire whenever they can. This simplifies our analysis. Even if shocks could be negative, the stability of the colour-blind equilibrium only depends on the slope of the best-response curve at zero. However, in this case, there might be stable equilibria different from those described in Theorem 1. In particular, it is possible that white employers prefer to hire black workers and

vice versa. That is, a social colour of an agent is more valuable if it is different from his physical colour.[13]

Social colour.– The social colour of an employer who does not hire becomes his physical colour conditional on the social colour changing. This can be motivated by assuming that if a white agent refuses to hire a black employee despite the positive surplus he will be considered loyal to other whites and hostile to blacks. However, the main reason for this assumption is that it enabled us to give a two-dimensional graphical representation of our problem. Recall that a consequence of this assumption is that the best response cutoff of an employer does not depend on his social colour (see Lemma 1 and (3)).

We assume that the social colour is a binary information and its evolution is only determined by the type of the worker and the physical colour of the employer. This is much in the spirit of Kandori (1992). One can model the information an employer observes about a worker in more complicated ways. For example, an employer might randomly draw a sample of the physical colour of the agents in the history of the worker. Then the type of the worker would be his physical colour and his full history. Such modelling would lead to a complex type space but does not alter our main result regarding the instability of the colour-bling equilibrium.

It is easy to construct social colours different from ours which do no lead to discrimination. For example, if this colour is not informative about past decisions then the colour-blind equilibrium is stable and unique. We have not characterized those processes which necessarily lead to discrimination.

More attributes and social colours.– In reality, there are more than one observable physical attributes of individuals. It is also possible that there are several labels attached to an individual conditional on his history. Of course, agents might condition their actions on these multi-dimensional attributes and labels. We emphasize, however, that as long as one of the dimensions of the label evolves as our social colour does, a version of Theorem 2 is still valid. That is, provided that λ is large, the colour-blind equilibrium is unstable and there exist stable equilibria. In other words, no matter how many attributes and labels are observable, adding the social colour destabilizes the colour-blind equilibrium. In this sense, our results are robust to more complicated information structures.

Constant wage.– Workers receive a constant wage, M, regardless of their types, the types of their employers and the profitability of the partnership. Therefore, any inefficiency due to discrimination is in the form of suboptimal unemployment decisions. In particular, an agent against whom others discriminate is only worse off because he is not employed frequently enough. It would be interesting to allow wages to be endogenous and analyze wage differentials due to racial discrimination. Unfortunately, it is not entirely clear how endogenous wages affect our main results. The problem is that if a black worker is willing to get a paycut in order to be employed by a white employer then there will be more white employers who employ black workers. This would

[13] These results are available upon request.

increase the number of white agents with black social colour, which in turn, makes it less costly for a white agent to have black social colour. This would make it less likely that discrimination arises in equilibrium. A potential solution for this problem is to allow the social colour to change as a function of the wage offered to a worker. The lower the wage of a black worker is the more likely it is that the employer's social colour becomes white.

We are currently working on models where wages are set endogenously. Preliminary results suggest that as long as the wage of a worker cannot fall to zero, the main results of our paper remain valid. There are various theories of wage determination, like efficiency wages and moral hazard problems, that lead to strictly positive wages even if the outside option of a worker is zero.

6.3 Empirical Implications

Our theory is different from taste-based and statistical discrimination theories because agents have no intrinsic preferences for interacting with others of the same colour, and their skin-colour provides no signal about their productivity. Next we discuss empirical predictions of our model which are different from that of these other two theories.

It is not hard to construct models of both statistical and taste-based theories which generate the same comparative static results as ours. The key observation is that past history of a worker affects the likelihood of him being hired. A white employer, for example, has a larger cutoff against a black worker than against a white one. As a result, the profit of a white employer who hires black workers is higher than that of those who hire white workers. This would also be true if discrimination was taste-based, and one can imagine a variation of statistical discrimination which also generates this result. However, a white employer also uses a larger cutoff against other whites with black social colour than against whites with white social colour. Therefore, white employers hiring white workers with black social colours earn more than those who hire whites with white social colours. Note that the social colour of an agent is more likely to be black if there were more blacks in his past history. Hence, our model predicts that the profit of a white employer from hiring a white worker is stochastically increasing in the number of black agents in the history of the worker.

In addition, agents would never discriminate in our model if they knew that their interactions were not observed. Our theory predicts that as the interactions become harder to observe it becomes less likely that discrimination arises. Therefore, people are more likely to discriminate in smaller communities, like villages, where people better observe the actions of others than in larger communities, such as large cities, where individuals have less information about each others. This is in sharp contrast to the predictions of the other two theories.

A notable feature of our model is that there exist stable equilibria where each race discriminates against the other one. This is inconsistent with the theory of statistical discrimination, according to which, the hiring decision of an employer should not depend on his own skin-colour.

7 Conclusion

This paper put forward a new theory of racial discrimination. Individuals discriminate against others because they do not want to be associated to the other race. Although the information about each others' association is not payoff-relevant, it plays a major role in determining the behavior of economic agents. Indeed, we showed that every stable equilibrium must involve discrimination in some environments.

Our model does not explain why agents use skin colour as a basis of discrimination as opposed to other observable physical attributes. People differ in height, weight, eye-colour, and along many other dimensions. A potential explanation might be that members of a family, or a community are more likely to have the same skin colour than the same height or weight. For a tall individual it would be more costly to discriminate against short ones if many of his relatives are short. Recall that a white agent discriminates against those who are associated to blacks in our model because he is afraid of those whites who are more associated to whites. Since individuals are necessarily associated to short and tall individuals, these attributes cannot be used to sustain discrimination. Another reason for using skin colour is that it easier to observe that other attributes such as the eye-colour.

We assumed throughout this paper that the surplus generated by a partnership is divided among the worker and employer exogenously. We have excluded the possibility that discrimination results in different wages conditional on employment. Perhaps the most important elaboration of our model would be to allow wages and profits to be determined endogenously.

We have not discussed policy in this paper. Recall that a white employer discriminates against black workers because he is afraid of not being hired by white employers with white social colour in the future. Hence, a policy intervention which would increase the fraction of those whose social colour is different from his physical colour reduces the incentive to discriminate. It is clear that subsidizing employers who hire workers of different physical colour will increase the fraction of those whose physical and social colour are different. This, in turn, decreases the fraction of those who have the same physical and social colours and reduce the incentive to discriminate. Such subsidies must be paid from taxes which might alter the incentives to produce. In order to discuss policy in a meaningful way, one must properly model production and the worker's incentives.

8 Appendix

8.1 Proof of the Lemmas

Proof of Lemma 3. (i) Notice that $b^c\left(x^c, x^{-c}\right) = \widetilde{b}^c\left(x^c, x^{-c}\right)$ whenever $b^c\left(x^c, x^{-c}\right) > 0$. Hence, it is enough to show that $\widetilde{b}^c$ is concave in x^c. By (9)

$$\frac{\partial \widetilde{b}^c\left(x^c, x^{-c}\right)}{\partial x^c} = K\mu_c\left(1-\alpha\right) g\left(\left(1-\alpha\right) x^c\right),$$

where $g(x) = \lambda e^{-\lambda x}$ for all $x \geq 0$. Therefore, this partial derivative is positive and decreasing.

(ii) By (9),

$$\widetilde{b}^c\left(0, x^{-c}\right) = K\left[\mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)\right] \leq 0,$$

because $G\left(\alpha \overline{x}^{-c}\right) - G\left(\overline{x}^{-c}\right) \leq 0$. Hence, (4) implies that $b^c\left(0, x^{-c}\right) = 0$.

(iii) Notice that $\lim_{x^{-c}\to\infty} G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right) = 0$. Therefore, by (9) and (4),

$$b^c\left(x^c, 0\right) = \lim_{x^{-c}\to\infty} b^c\left(x^c, x^{-c}\right) = K\mu_c G\left(\left(1-\alpha\right)x^c\right).$$

Finally, since $G\left(\alpha \overline{x}^{-c}\right) - G\left(\overline{x}^{-c}\right) \leq 0$,

$$b^c\left(x^c, 0\right) \geq b^c\left(x^c, \overline{x}^{-c}\right).$$

■

Proof of Lemma 4. Notice that if $x^{-c} = 0$ then $\widetilde{b}^c\left(x^c, x^{-c}\right) \geq 0$, and by (4), $\widetilde{b}^c\left(x^c, 0\right) = b^c\left(x^c, 0\right)$. We have showed in the proof of part (i) of Lemma 3 that

$$\frac{\partial b^c\left(x^c, 0\right)}{\partial x^c} = K\mu_c\left(1-\alpha\right)\lambda e^{-\lambda(1-\alpha)x^c}.$$

This derivative is $K\mu_c\left(1-\alpha\right)\lambda$ at $x^c = 0$, and converges to zero as x^c goes to infinity.

(i) If $\lambda > \lambda_0$, $K\mu_c\left(1-\alpha\right)\lambda > 1$ and therefore $b^c\left(x^c, 0\right) > x^c$ around zero. Since the curve $b^c\left(x^c, 0\right)$ is concave (part (i) of Lemma 3) and its derivative goes to zero as x^c goes to infinity, there exists a unique $x^c > 0$ such that $b^c\left(x^c, 0\right) = x^c$.

(ii) If $\lambda \leq \lambda_0$, $K\mu_c\left(1-\alpha\right)\lambda \leq 1$. Since the curve $b^c\left(x^c, 0\right)$ is concave (part (i) of Lemma 3) $b^c\left(x^c, 0\right) < x^c$ for all $x^c > 0$. ■

Proof of Lemma 5. For each x^{-c} consider the following function of x^c:

$$\begin{aligned} B^{x^{-c}}\left(x^c\right) &= \widetilde{b}^c\left(x^c, x^{-c}\right) - x^c \\ &= K\mu_c G\left(\left(1-\alpha\right)x^c\right) - x^c + K\mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right) \end{aligned}$$

Observe that, by (4), $b^c\left(x^c, x^{-c}\right) = x^c$ if and only if x^c is a root of $B^{x^{-c}}$. First, we establish some properties of $B^{x^{-c}}$.

(1) The function $B^{x^{-c}}$ is strictly concave . This follows from the proof of part (i) of Lemma 3.

(2) $dB^{x^{-c}}/dx^c\Big|_{x_c=0} > 0$. It follows from the proof of part (i) of Lemma 4.

(3) $\lim_{x^c\to\infty} B^{x^{-c}}\left(x^c\right) = -\infty$. This is because G is a CDF and hence, $\widetilde{b}^c\left(x^c, x^{-c}\right) \leq K$.

(4) $\lim_{x^c\to 0} B^{x^{-c}}\left(x^c\right) < 0$. This is because $G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)$ is negative.

(5) Generically, $B^{x^{-c}}$ has either zero or two roots. This follows from (1)-(4).[14]

Notice that the part $B^{x^{-c}}\left(x^c\right)$ which depends on x^{-c} is additively separable from the rest. Hence, any change in x^{-c} results in a vertical shift of this curve. Whether there are two or zero

[14] The non-generic case is when this function is tangent to the constant zero line. There can be at most one such an x^{-c}.

intersections depends on the size of this shift. Let $H\left(x^{-c}\right)=K\mu_{-c}\left(G\left(\alpha x^{-c}\right)-G\left(x^{-c}\right)\right)$. Next we establish some properties of this function.[15]

(6) $H\left(x^{-c}\right)$ is strictly negative.

(7) $\lim_{x^{-c}\to\infty}H\left(x^{-c}\right)=\lim_{x^{-c}\to 0}H\left(x^{-c}\right)=0$.

(8) $\arg\min H\left(x^{-c}\right)=\left(-\log\alpha\right)/\left[\lambda\left(1-\alpha\right)\right]=\widehat{x}^{-c}$.

(9) H is strictly decreasing on $\left(0,\widehat{x}^{-c}\right)$ and strictly increasing on $\left(\widehat{x}^{-c},\infty\right)$.

We are ready to prove the two parts of the lemma.

(i) Suppose that $\max B^{\widehat{x}^{-c}}\left(x^{c}\right)>0$. This, together with (8), implies that $\max B^{\widehat{x}^{-c}}\left(x^{c}\right)>0$ for all $x^{-c}>0$. By (3) the value of $B^{x^{-c}}$ is sometimes negative. The intermediate Value Theorem implies that $B^{x^{-c}}$ has at least one root. Hence, by (5), has exactly two roots.

(ii) Suppose that $\max B^{\overline{x}^{-c}}\left(x^{c}\right)<0$. This means that there are values of x^{-c} for which $B^{x^{-c}}$ is always negative. (9) implies that the set of such $x^{-c}s$ is an interval. Let us denote this interval by $\left(\underline{x}^{-c},\overline{x}^{-c}\right)$. From (2) and (7) it follows that $\underline{x}^{-c}>0$ and $\overline{x}^{-c}<\infty$. The argument establishing that $B^{x^{-c}}$ has two roots whenever $x^{-c}\in\mathbb{R}_{+}\backslash\left[\underline{x}^{-c},\overline{x}^{-c}\right]$ is analogous to the argument in part (i). From (1) and (9) it follows that

$$\lim_{x^{-c}\to\underline{x}^{-c}}x_1^c\left(x^{-c}\right)=\lim_{x^{-c}\to\overline{x}^{-c}}x_1^c\left(x^{-c}\right)=\lim_{x^{-c}\to\overline{x}^{-c}}x_2^c\left(x^{-c}\right)\lim_{x^{-c}\to\underline{x}^{-c}}x_2^c\left(x^{-c}\right).$$

It remains to show that $x_1^c\left(x^{-c}\right)$ is increasing first, then is decreasing, and $x_2^c\left(x^{-c}\right)$ is decreasing first, then is increasing. On the interval $\left(0,\widehat{x}^{-c}\right)$ an increase in x^{-c} results a downwards shift of $B^{x^{-c}}$ (see (9)). Hence, by (1), $x_1^c\left(x^{-c}\right)$ is increasing and $x_2^c\left(x^{-c}\right)$ is decreasing on this interval. On the interval $\mathbb{R}_{+}\backslash\left(0,\widehat{x}^{-c}\right)$ an increase in x^{-c} results an upward shift of $B^{x^{-c}}$ (see (9)). Hence, by (1), $x_1^c\left(x^{-c}\right)$ is decreasing and $x_2^c\left(x^{-c}\right)$ is increasing on this interval. Finally, it follows from (2) and (7) that $\lim_{x^{-c}\to 0}x_1^c\left(x^{-c}\right)=0$. ■

8.2 Proof of Proposition 1

Before we proceed with the proof of Proposition 1 we prove a few Lemmas about the equilibrium cutoffs. For convenience we introduce a few new notations. We shall denote $\min\left\{\mu_b,\mu_w\right\}$ by $\mu_{\min}$. In addition, we define two constants

$$\begin{aligned}\psi_0&=\frac{1}{\frac{1}{4}K\alpha\left(1-\alpha\right)\mu_{\min}},\\ \psi_1&=K\mu_{\min}\frac{1}{2}\left(1-2^{-\alpha}\right)\left(1-2^{-(1-\alpha)}\right).\end{aligned}\tag{13}$$

In the proofs of the lemmas we often use the inequality stated in the next

Lemma 6 *For all* $\xi\leq\lg 2$

$$1-e^{-\xi}\geq\frac{1}{2}\xi.\tag{14}$$

[15] These properties are streightforward consequences of the assumption that $G\left(s\right)=1-e^{\lambda s}$.

In what follows (x^c, x^{-c}) denotes an equilibrium cutoff profile.

Lemma 7 *There exists a λ_0 such that for all $\lambda \geq \lambda_0$ either* $\max\{x^c, x^{-c}\} \leq \psi_0 \lambda^{-2}$ *or* $\max\{x^c, x^{-c}\} \geq \psi_1$.

Proof. First, suppose that both cutoffs are strictly positive, that is, $x^b, x^w > 0$. Then, by (8),

$$\begin{aligned} x^b + x^w &= \sum\nolimits_{c\in\{b.w\}} K\left[\mu_c G\left(\left(1-\alpha\right)x^c\right) + \mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)\right] \\ &= \sum\nolimits_{c\in\{b.w\}} K\mu_c\left[G\left(\left(1-\alpha\right)x^c\right) + G\left(\alpha x^c\right) - G\left(x^c\right)\right] \\ &= \sum\nolimits_{c\in\{b,w\}} K\mu_c\left(1-e^{-\alpha\lambda x^c}\right)\left(1-e^{-(1-\alpha)\lambda x^c}\right), \end{aligned} \tag{15}$$

where the first equality follows from rearranging the terms corresponding to the same colour and the second one from $G(x) = 1 - e^{-x}$ and

$$\left[1-e^{-\alpha\lambda x^c}\right] + \left[1-e^{-(1-\alpha)\lambda x^c}\right] + \left[1-e^{-\lambda x^c}\right] = \left(1-e^{-\alpha\lambda x^c}\right)\left(1-e^{-(1-\alpha)\lambda x^c}\right).$$

We consider two cases. If $\max\left\{x^b, x^w\right\} \geq (\lg 2)/\lambda$, then from the previous equality it follows that

$$\begin{aligned} x^b + x^w &\geq K\mu_{\min}\left(1-e^{-\alpha\lg 2}\right)\left(1-e^{-(1-\alpha)\lg 2}\right) \\ &= K\mu_{\min}\left(1-2^{-\alpha}\right)\left(1-2^{-(1-\alpha)}\right) = 2\psi_1. \end{aligned}$$

Since $\max\left\{x^b, x^w\right\} \geq x^b + x^w$, the previous inequality chain implies $\max\left\{x^b, x^w\right\} \geq \psi_1$. If $\max\left\{x^b, x^w\right\} \leq (\log 2)/\lambda$, then, by Lemma 6,

$$1-e^{-\alpha\lambda x^c} \geq \frac{1}{2}\alpha\lambda x^c \text{ and } 1-e^{-(1-\alpha)\lambda x^c} \geq \frac{1}{2}\left(1-\alpha\right)\lambda x^c \tag{16}$$

for each $c \in \{b, w\}$. Equations (15) and inequalities (16) imply that

$$\begin{aligned} \max\left\{x^b, x^w\right\} &\geq \sum\nolimits_{c\in\{b,w\}} \frac{1}{4}K\alpha\left(1-\alpha\right)\mu_c\lambda^2\left(x^c\right)^2 \\ &\geq \frac{1}{4}K\alpha\left(1-\alpha\right)\mu_{\min}\lambda^2\left[\left(x^c\right)^2 + \left(x^{-c}\right)^2\right] \geq \frac{1}{\psi_0}\lambda^2\left(\max\left\{x^b, x^w\right\}\right)^2. \end{aligned}$$

Hence, $\max\left\{x^b, x^w\right\} \leq \psi_0/\left(\lambda^2\right)$.

Second, suppose that one of the cutoffs is zero, and without loss of generality assume that $x^b = 0$ and, hence, $\max\left\{x^b, x^w\right\} = x^w$. Then, by (8),

$$x^w = K\mu_w\left(1-e^{-(1-\alpha)\lambda x^w}\right).$$

If $x^w \geq (\log 2)/\lambda$ then

$$x^w \geq K\mu_w\left(1-e^{-(1-\alpha)\lg 2}\right) \geq \psi_1.$$

If $x^w \leq (\log 2)/\lambda$ then, by Lemma 6,

$$x^w \geq 2K\mu_w\left(1-\alpha\right)\lambda x^w.$$

If $\lambda > 1/\left(2K\mu_w\left(1-\alpha\right)\right)$ then the previous inequality implies that $x^w \leq 0$ and hence, $x^w < \psi_0\lambda^{-2}$.

■

Lemma 8 *There exists a λ_0 such that if $\lambda \geq \lambda_0$ and $x^c \geq \psi_1$ then either $x^{-c} \leq \psi_0/(\lambda)^2$ or $x^{-c} \geq \psi_1/2$.*

Proof. Suppose that $x^c \geq \psi_1$. If $x^{-c} < 0$ then $x^{-c} \geq \psi_1/2$. Suppose now that $x^{-c} > 0$. Then

$$\begin{aligned} x^{-c} &= K\mu_{-c} G\left((1-\alpha)x^{-c}\right) + K\mu_c\left(G(\alpha x^c) - G(x^c)\right) \\ &\geq K\mu_{-c}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) - K\mu_c e^{-\lambda\alpha\psi_1}, \end{aligned} \tag{17}$$

where the equality is just (8) and the inequality follows from $x^c \geq \psi_1$. We consider two cases.

Case 1: $x^{-c} \geq (\lg 2)/\lambda$. If λ is large enough so that $K\mu_c e^{-\lambda\alpha\psi_1} \leq \frac{1}{2}\psi_1$,

$$\begin{aligned} K\mu_{-c}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) - K\mu_c e^{-\lambda\alpha\psi_1} &\geq K\mu_{-c}\left(1 - e^{-(1-\alpha)\lg 2}\right) - \frac{1}{2}\psi_1 \\ &\geq K\mu_{-c}\left(1 - 2^{-(1-\alpha)}\right) - \frac{1}{2}\psi_1 \geq \frac{1}{2}\psi_1, \end{aligned}$$

where the last equality follows from $\psi_1 \leq K\mu_{-c}\left(1 - 2^{-(1-\alpha)}\right)$. The previous inequality chain and (17) imply $x^{-c} \geq \frac{1}{2}\psi_1$.

Case 2: $x^{-c} < (\lg 2)/\lambda$. Then, by Lemma 6,

$$1 - e^{-\lambda(1-\alpha)x^{-c}} \geq \frac{1}{2}(1-\alpha)\lambda x^{-c}. \tag{18}$$

If λ is large enough so that $K\mu_{\max} e^{-\lambda\alpha\psi_1} \leq \psi_0/(\lambda)^2$, the previous inequality implies that

$$K\mu_{-c}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) - K\mu_c e^{-\lambda\alpha\psi_1} \geq K\mu_{\min}\frac{1}{2}(1-\alpha)\lambda x^{-c} - \psi_0\lambda^{-2}.$$

This inequality and the inequality chain (17) yields

$$\left(K\mu_{\min}\frac{1}{2}(1-\alpha)\lambda - 1\right)x^{-c} \leq \frac{\psi_0}{\lambda^2}.$$

If λ is large enough so that $K\mu_{\min}\frac{1}{2}(1-\alpha)\lambda - 1 > 1$ then $x^{-c} \leq \psi_0\lambda^{-2}$. ■

Recall that $x_{\max}$ is the largest possible cutoff which can be a best response to a cutoff profile and $x_{\max} = K\mu_c$.

Lemma 9 *For all $\varepsilon > 0$, there exists a λ_0, such that if $\lambda > \lambda_0$ and $x^c \geq \psi_1/2$ then either $x^{-c} \in \left(\psi_0/(\lambda)^2, \psi_1/2\right)$ or $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$.*

Proof. Suppose that $x^c \geq \psi_1/2$ and that $x^{-c} \in \left(\psi_0/(\lambda)^2, \psi_1/2\right)$. It is enough to show that for all $\varepsilon > 0$ there is a λ_0 such that if $\lambda > \lambda_0$ then whenever $x^c \geq \psi_1/2$ and $x^{-c} \in \left(\psi_0/(\lambda)^2, \psi_1/2\right)$ the cutoff x^c must be in the set $(x^c_{\max} - \varepsilon, x^c_{\max})$.

Notice that from (8) and $x_{\max} = K\mu_c$ it follows that

$$\begin{aligned} x^c_{\max} - x^c &= K\mu_c - \left[K\mu_c\left(1 - e^{-\lambda(1-\alpha)x^c}\right) + K\mu_{-c}\left(1 - e^{-\lambda x^{-c}} - 1 + e^{-\lambda\alpha x^{-c}}\right)\right] \\ &= K\mu_c e^{-\lambda(1-\alpha)x^c} - K\mu_{-c}\left(e^{-\lambda x^{-c}} - e^{-\lambda\alpha x^{-c}}\right) \\ &= K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c} e^{-\lambda\alpha x^{-c}}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right). \end{aligned} \tag{19}$$

Case 1: $x^{-c} \geq \psi_1/2$. Then

$$\begin{aligned} K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c} e^{-\lambda\alpha x^{-c}} \left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) &\leq K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c} e^{-\lambda\alpha x^{-c}} \\ &\leq K\mu_c e^{-\frac{1}{2}\lambda(1-\alpha)\psi_1} + K\mu_{-c} e^{-\frac{1}{2}\lambda\alpha\psi_1}, \end{aligned}$$

where the first inequality follows from $1 - e^{-(1-\alpha)\lambda x^{-c}} \leq 1$ and the second one from $x^{-c}, x^c \geq \psi_1/2$. This inequality chain and (19) imply that

$$x^c_{\max} - x^c \leq K\mu_c e^{-\frac{1}{2}\lambda(1-\alpha)\psi_1} + K\mu_{-c} e^{-\frac{1}{2}\lambda\alpha\psi_1}.$$

Notice that for each ε there is a λ_0 such that if $\lambda > \lambda_0$ the right-hand-side of this inequality is smaller than ε and, hence, $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$.

Case 2: If $x^{-c} \leq \psi_0 / (\lambda)^2$, then,

$$\begin{aligned} K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c} e^{-\lambda\alpha x^{-c}} \left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) &\leq K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c} \left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) \\ &\leq K\mu_c e^{-\frac{1}{2}\lambda(1-\alpha)\psi_1} + K\mu_{-c} \left(1 - e^{\frac{-\psi_0(1-\alpha)}{\lambda}}\right), \end{aligned}$$

where the first inequality follows from $e^{-\lambda\alpha x^{-c}} \leq 1$ and the second one from $x^c \geq \psi_1/2$ and $x^{-c} \leq \psi_0 / (\lambda)^2$. This inequality chain and (19) imply that

$$x^c_{\max} - x^c \leq K\mu_c e^{-\frac{1}{2}\lambda(1-\alpha)\psi_1} + K\mu_{-c} \left(1 - e^{\frac{-\psi_0(1-\alpha)}{\lambda}}\right).$$

Observe that as λ goes to infinity both $K\mu_c e^{-(1/2)\lambda(1-\alpha)\psi_1}$ and $1 - e^{-\psi_0(1-\alpha)/\lambda}$ converge to zero. Therefore, for each ε there is a λ_0 such that if $\lambda > \lambda_0$ the right-hand-side of this inequality is smaller than ε and $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$. ■

We are ready to prove Proposition 1. By Lemma 7, we have to consider two cases: either $\max\{x^c, x^{-c}\} \leq \psi_0 \lambda^{-2}$ or $\max\{x^c, x^{-c}\} \geq \psi_1$.

Case 1: $\max\{x^c, x^{-c}\} \leq \psi_0 \lambda^{-2}$. Note that for each ε there is a λ_0 such that for all $\lambda > \lambda_0$ the term $\psi_0 \lambda^{-2}$ is strictly smaller than ε, and hence, $x^b, x^w < \varepsilon$. Therefore, either (i) or (ii) holds in the statement of Proposition 1.

Case 2: $\max\{x^c, x^{-c}\} \geq \psi_1$. Without loss of generality assume that $\max\{x^c, x^{-c}\} = x^c$. By Lemma 8, we have to consider only the following two subcases: either $x^{-c} \leq \psi_0 / (\lambda)^2$, or $x^{-c} \geq (1/2)\psi_1$.

Case 2.a: $x^{-c} \leq \psi_0 / (\lambda)^2$. Then for each ε there is a λ_0 such that if $\lambda \geq \lambda_0$ then $x^{-c} \leq \varepsilon$, and by Lemma 9, $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$. In this case (i) holds for x^{-c} and (iii) holds for x^c.

Case 2.b: $x^{-c} \geq (1/2)\psi_1$. Since $x^c \geq \psi_1 > (1/2)\psi_1$, Lemma 9 implies that $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$ for $c \in \{b, w\}$. Therefore, (iii) of the statement of Proposition 1 holds for $c \in \{b, w\}$.

8.3 Proof of Theorem 1

By Proposition 1, for all ε there exists a λ_0 such that any equilibria can be classifies into one of the cases described by the following table.

	x^c	x^{-c}
Case 1	0	0
Case 2	0	$\in (0, \varepsilon)$
Case 3	0	$\in (x^{-c}_{\max} - \varepsilon, x^{-c}_{\max})$
Case 4	$\in (0, \varepsilon)$	$\in (0, \varepsilon)$
Case 5	$\in (0, \varepsilon)$	$\in (x^{-c}_{\max} - \varepsilon, x^{-c}_{\max})$
Case 6	$\in (x^{-c}_{\max} - \varepsilon, x^{-c}_{\max})$	$\in (x^{-c}_{\max} - \varepsilon, x^{-c}_{\max})$

We shall consider each case separately. We show that equilibria described by Case 2 do not exist and equilibria corresponding to Cases 1, 4, and 5 are unstable. Finally, we prove that equilibria corresponding to Cases 3 and 6 are unique and stable. Note that this accomplishes the proof of Theorem 1. In what follows, we use the notations introduces in the last subsection, see (13).

Case 2.

In order to show that there does not exists an equilibrium described by Case 2, it is enough to prove that if λ is large enough and $x^c = 0$ then $x^{-c} = 0$ or $x^{-c} \geq \psi_1$. By Lemma 7, $x^{-c} \geq \psi_1$ or $x^{-c} \leq \psi_0 / (\lambda)^2$. If $x^{-c} \geq \psi_1$, we are done. It remains to be shown that $x^{-c} \leq \psi_0 / (\lambda)^2$ implies $x^{-c} = 0$. We prove it by contradiction, and assume that $x^{-c} \in (0, \psi_0 / (\lambda)^2]$. Then,

$$x^{-c} = K\mu_{-c} G\left((1-\alpha) x^{-c}\right) = K\mu_{-c}\left(1 - e^{-\lambda(1-\alpha)x^{-c}}\right) \geq \frac{1}{2} K\mu_{-c}\lambda (1-\alpha) x^{-c} > x^{-c},$$

where the first equality is just (8) with $x^c = 0$, the first inequality follows from Lemma 6, and the second one from λ being large. Note that the previous inequality chain cannot hold, hence, $x^{-c} = 0$.

Case 1.

We show that the equilibrium cutoff profile $(0, 0)$ is unstable. By Definition 1, it is enough to show that there exists a distribution of cutoff profiles *nearby* $(0, 0)$ such that the best-response dynamics does not converge to $(0, 0)$. To this end, choose X^c and X^{-c} to be deterministic variables such that $X^{-c} = 0$ and $X^c = \delta$, where $\delta \in (0, \lg 2 / [\lambda (1-\alpha)])$. Let the best response of an agent with colour c at time t denoted by x^c_t if the initial distribution of cutoffs is (X^c, X^{-c}). Equations (8) and (12) imply that $x^{-c}_t = 0$ for all t. However, we show that x^c_t does not converge to 0 for sufficiently large λ. Since $x^c_0 > 0$, it is enough to prove that $dx^c_t/dt > 0$ whenever x^c_t is small but positive. Suppose that $x^c_t \in (0, \lg 2 / [\lambda (1-\alpha)])$. Then

$$\begin{aligned} \frac{dx^c_t}{dt} &= K\mu_c \left(1 - e^{-\lambda(1-\alpha)x^c_t}\right) - x^c_t \\ &\geq \lambda \frac{1}{2} K\mu_c (1-\alpha) x^c_t - x^c_t = \left(\lambda \frac{1}{2} K\mu_c (1-\alpha) - 1\right) x^c_t, \end{aligned}$$

where the first equality is just (12) with $x_t^{-c}=0$ and the inequality follows from $x_t^c \in (0, \lg 2/[\lambda(1-\alpha)])$ and Lemma 6. If λ is large enough then $\lambda K \mu_c (1-\alpha)/2 > 1$, and hence, $dx_t^c/dt > 0$.

Cases 4 and 5.

Using the equation describing the best-response dynamics, (12), we construct the Jacobian matrix corresponding to the dynamic system $\left(x_t^c, x_t^{-c}\right)$:

$$J\left(x_t^c, x_t^{-c}\right) = \begin{bmatrix} \frac{d\tilde{b}^c\left(x_t^c, x_t^{-c}\right)}{dx_t^c} - 1, & \frac{d\tilde{b}^c\left(x_t^c, x_t^{-c}\right)}{dx_t^{-c}}, \\ \frac{d\tilde{b}^{-c}\left(x_t^c, x_t^{-c}\right)}{dx_t^c}, & \frac{d\tilde{b}^{-c}\left(x_t^c, x_t^{-c}\right)}{dx_t^{-c}} - 1 \end{bmatrix}. \tag{20}$$

where all the derivatives are taken at $\left(x_0^c, x_0^{-c}\right)$. Since in Cases 4 and 5 $x_0^c, x_0^{-c} > 0$, the Hartman-Grobman Theorem implies that $\left(x_0^c, x_0^{-c}\right)$ is not a stable equilibrium if an eigenvalue of $J\left(x_0^c, x_0^{-c}\right)$ has a positive real part. It is well-known that if $\operatorname{tr} J\left(x_0^c, x_0^{-c}\right) > 0$ or $\det D\left(x_0^c, x_0^{-c}\right) < 0$, then the real part of at least of the eigenvalues is positive. Therefore, in order to establish that $\left(x_0^c, x_0^{-c}\right)$ is unstable it is enough to show that $\operatorname{tr} J\left(x_0^c, x_0^{-c}\right) > 0$.

In Case 4, Proposition 1 and Lemma 7 imply that $x^c \in \left(0, \psi_0/(\lambda)^2\right)$ if λ is large enough. In Case 5, Proposition 1 and Lemma 8 imply that $x^c \in \left(0, \psi_0/(\lambda)^2\right)$ if λ is large enough. Also notice that

$$\frac{d\tilde{b}^{-c}(x^{-c}, x^c)}{dx^{-c}} = K\mu_{-c}\lambda(1-\alpha)e^{-\lambda(1-\alpha)x^{-c}} \geq 0,$$

and for sufficiently large λ,

$$\begin{aligned} \frac{d\tilde{b}^c(x^c, x^{-c})}{dx^c} &= K\mu_c\lambda(1-\alpha)e^{-\lambda(1-\alpha)x^c} \geq K\mu_c\lambda(1-\alpha)\left(1 - e^{-(1-\alpha)\psi_0/\lambda}\right) \\ &\geq \frac{1}{2}K\mu_c\lambda(1-\alpha), \end{aligned}$$

where the first inequality follows from $x^c < \psi_0/(\lambda)^2$ and the second one from $e^{-(1-\alpha)\psi_0/\lambda} < 1/2$ if λ is large. Therefore, if λ is large enough,

$$\begin{aligned} \operatorname{tr} J\left(x_0^c, x_0^{-c}\right) &= \frac{d\tilde{b}^c\left(x_0^c, x_0^{-c}\right)}{dx_0^c} - 1 + \frac{d\tilde{b}^{-c}\left(x_0^c, x_0^{-c}\right)}{dx_0^{-c}} - 1 \\ &\geq \frac{1}{2}\lambda K \mu_{\min}(1-\alpha) - 2 > 0. \end{aligned}$$

Case 3.

Remark 2 established that, if λ is large, the equilibrium exists and is unique in this case. It remains to show that this equilibrium is stable. Notice that this equilibrium corresponds to the intersection of the x_2^{-c} and x_0^c curves, that is, $\left(0, x_2^{-c}(0)\right)$. Since the curve x_2^{-c} is continuous, there exist δ_1 and δ_2 such that if $x^c < \delta_1$ then $|x^{-c} - x_2^{-c}(x^c)| < \delta_2$. In addition, we established in Section 5 that if δ_1 and δ_2 is small enough, $x^c < \delta_1$ and $|x^{-c} - x_2^{-c}(0)| < \delta_2$ then

$$\widetilde{b}^c\left(x^c, x^{-c}\right) - x^c < 0 \text{ and } \widetilde{b}^{-c}\left(x^{-c}, x^c\right) - x^{-c} \begin{cases} > 0 & \text{if } x^{-c} < x_2{}^{-c}(x^c), \\ < 0 & \text{if } x^{-c} > x_2{}^{-c}(x^c), \\ = 0 & \text{if } x^{-c} = x_2{}^{-c}(x^c). \end{cases} \tag{21}$$

Let δ be so small that for any cutoff distribution (X^c, X^{-c}), if $|X^c| < \delta$ and $|X^{-c} - x_2^{-c}(0)| < \delta$ almost surely then the initial best-response cutoff profile, $\left(x_0^c, x_0^{-c}\right)$, satisfy $x^c < \delta_1$ and $|x^{-c} - x_2^{-c}(0)| < \delta_2$. Then (21) implies that $\left(x_t^c, x_t^{-c}\right)$ is in the rectangle

$$\left\{\left(x^c, x^{-c}\right) : x^c \in (0, \delta_1),\, |x^{-c} - x_2^{-c}(0)| < \delta_2\right\}$$

for all t.[16] Therefore, $\lim_{t\to\infty} x_t^c = 0$ by (21). This, together with (21), implies $\lim x_t^{-c} = x_2^{-c}(0)$.

Case 6.

First, we show that if this equilibrium exists it is stable. Recall the matrix introduced in Cases 4 and 5, $J\left(x_0^c, x_0^{-c}\right)$. Since $x_0^c, x_0^{-c} > 0$, we can apply the Hartman-Grobman Theorem which implies that $\left(x_0^c, x_0^{-c}\right)$ is a stable equilibrium if all eigenvalues of $J\left(x_0^c, x_0^{-c}\right)$ have negative real parts. It is well-known that if $\operatorname{tr} D\left(x_0^b, x_0^w\right) < 0$ and $\det D\left(x_0^b, x_0^w\right) > 0$ then the eigenvalues indeed have negative real parts. In this case, if λ is large enough then $x^b, x^w > x_{\max} - \varepsilon > \psi_1/2$. In addition, for all $\delta > 0$ there is a λ_0 such that if $\lambda > \lambda_0$,

$$\frac{d\tilde{b}^c\left(x^c, x^{-c}\right)}{dx^c} - 1 = \lambda K \mu_c (1-\alpha) e^{-\lambda(1-\alpha)x^c} - 1 \in \left(-1, -1 + \frac{\delta}{2}\right), \tag{22}$$

and

$$\begin{aligned}\frac{d\tilde{b}^c\left(x^c, x^{-c}\right)}{dx^{-c}} &= \lambda K \mu_{-c}\left(\alpha e^{-\lambda\alpha x^{-c}} - e^{-\lambda x^{-c}}\right) \\ &= \lambda K \mu_{-c} e^{-\lambda\alpha x^{-c}}\left(\alpha - e^{-\lambda(1-\alpha)x^{-c}}\right) \in \left(0, \frac{\delta}{2}\right).\end{aligned}$$

Thus,

$$\operatorname{tr} D\left(x_0^c, x_0^{-c}\right) < -2 + \delta < 0 \text{ and } \det D\left(x_0^c, x_0^{-c}\right) > 1 - \delta^2 > 0.$$

In order to show the existence of an equilibrium in this case, we show that the curves x_2^c and x_2^{-c} are defined on $[\psi_1/2, \infty)$ and they intersect. By (22), $\tilde{b}^c\left(x^c, x^{-c}\right) - x^c$ is strictly decreasing in x^c on this interval. Since $\tilde{b}$ is bounded from above by $x_{\max}$, $\lim_{x^c\to\infty}\left[\tilde{b}^c\left(x^c, x^{-c}\right) - x^c\right] = -\infty$. In addition, Lemma 9 implies that, $\tilde{b}^c\left(x^c, x^{-c}\right) \geq x_{\max}^c - \varepsilon = K\mu_c - \varepsilon > \psi_1/2$ if $x^c, x^{-c} \in [\psi_1/2, \infty)$. Therefore, $\tilde{b}^c\left(x^c, x^{-c}\right) - x^c$ is strictly decreasing, positive at $x^c = \psi_1/2$, and becomes negative as x^c gets large whenever $x^{-c} \in [\psi_1/2, \infty)$. Therefore, for each $x^{-c} \in [\psi_1/2, \infty)$ there exists exactly one x^c such that $\tilde{b}^c\left(x^c, x^{-c}\right) = x^c$. We denotes this x^c by $x_2^c\left(x^{-c}\right)$ (see Lemma 5). Lemma 9 implies that $x_2^c\left(x^{-c}\right) \in [x_{\max}^c - \varepsilon, x_{\max}^c]$ for all $x^{-c} \geq \psi_1/2$. Since this argument holds for both c, the mapping $x_2^c \circ x_2^{-c} : \left[\frac{1}{2}\psi_1, \infty\right) \to [x_{\max}^c - \varepsilon, x_{\max}^c]$ is well-defined and clearly continuous. Therefore, there exists an $x_*^c \in [x_{\max}^c - \varepsilon, x_{\max}^c]$ such that

$$x_2^c\left(x_2^{-c}\left(x_*^c\right)\right) = x_*^c.$$

Define $x_*^{-c} = x_2^{-c}\left(x_*^c\right)$. Then, by (11), (x_*^c, x_*^{-c}) is an equilibrium cutoff profile.

[16] This is because $dx_t^c dt < 0$ whenever $x_t^c = \delta_1$, $dx_t^{-c}/dt < 0$ if $x_t^{-c} = x_2^{-c}(0) + \delta_2$ and $dx_t^{-c}/dt > 0$ if $x_t^{-c} = x_2^{-c}(0) - \delta_2$.

In order to show the uniqueness, consider the mapping $B : [\psi_1/2, \infty)^2 \to \mathbb{R}^2$ defined by

$$B\left(x^c, x^{-c}\right) = \left(\tilde{b}^c\left(x^c, x^{-c}\right) - x^c, \tilde{b}^{-c}\left(x^{-c}, x^c\right) - x^{-c}\right).$$

Note that (x^c, x^{-c}) is an equilibrium if and only if $B\left(x^c, x^{-c}\right) = (0, 0)$. Note that the Jacobian matrix of B is just $J\left(x^c, x^{-c}\right)$. We have concluded above that the determinant of this matrix is strictly positive on $x^c, x^{-c} \in [\psi_1/2, \infty)$. Therefore, B is an injection and there can only be at most one (x^c, x^{-c}) satisfying $B\left(x^c, x^{-c}\right) = (0, 0)$.

References

Alesina, A., and E. L. Ferrara (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43, 721–61.

Arrow, K. J. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter, and A. Rees, pp. 3–33. Princeton University Press.

Austen-Smith, D., and R. G. F. Jr. (2005): "An Economic Analysis of 'Acting White'.," *Quarterly Journal of Economics*, 120, 551–583.

Baccara, M. G., and L. Yariv (2008): "Similarity and Polarization in Groups," .

Becker, G. S. (1971): *The Economics of Discrimination*. University of Chicago Press, Chicago.

Coate, S., and G. C. Loury (1993): "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *American Economic Review*, 83(5), 1220–40.

Eeckhout, J. (2006): "Minorities and Endogenous Segregation," *Review of Economic Studies*, 254, 31–53.

Fang, H., and A. Moro (2010): "Theories of Statistical Discrimination and Affirmative Action: A Survey," in *Handbook of Social Economics, Vol.*, ed. by J. Benhabib, A. Bisin, and M. Jackson, p. ???

Lang, K., M. Manove, and W. T. Dickens (2005): "Racial Discrimination in Labor Markets with Posted Wage Offers," *American Economic Review*, 95(4), 1327–1340.

Mailath, G., L. Samuelson, and A. Shaked (2000): "Endogenuous Inequaluity in Integrated Labor Markets with Two-Sided Search," *American Economic Review*, 90, 46–72.

Mailath, G. J., and A. Postlewaite (2006): "Social Assets," *International Economic Review*, 47, 1057–1091.

Moro, A., and P. Norman (2004): "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory*, 114, 1–30.

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

Ridley, M. (1993): *The Red Queen: Sex and the Evolution of Human Nature*. Penguin, London.

Rosén, Å. (1997): "An Equilibrium Search-Matching Model of Discrimination," *European Economic Review*, 41, 1589–1613.

Schelling, T. S. (1971): "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1, 143–186.