Spontaneous Discrimination*

Marcin Pęski† and Balázs Szentes‡

November 30, 2012

Abstract

This paper considers a dynamic economy in which agents are repeatedly matched with one another and then decide whether or not to enter into profitable partnerships. Each agent has a physical colour and a *social colour*. The social colour of an agent acts as a signal, conveying information about the physical colour of agents in his partnership history. Before an agent makes a decision, he observes his match's physical and social colours. Neither the physical colour nor the the social colour is payoff-relevant. We identify environments where equilibria arise in which agents condition their decisions on the physical and social colours of their potential partners. That is, they discriminate.

1 Introduction

This paper proposes a new theory of racial discrimination. According to our theory, discrimination can arise *spontaneously* in the form of a social norm, without any intrinsic tastes for race or any differences between individuals. Individuals discriminate against the other race because they would otherwise face discrimination from their own race. In order for this behaviour to arise, the crucial requirement is that before one decides whether to interact with another individual, he observes some coarse information about the other's past social interactions, and specifically the race of those that other person has chosen to associate himself with directly or indirectly. Agents can then condition their decisions on this information. One consequence, and indeed the main result of this paper, is that even if each individual is basically tolerant of other races, agents might prefer to interact only with those of the same race, and might also avoid those who even indirectly associate themselves with the opposite race. In other words, being associated with one's own race becomes valuable through the equilibrium play.

In the specific model analysed in this paper, agents are repeatedly matched with one another. After being matched, each agent must decide whether or not to enter into a profitable relationship

*We have benefited from discussions with Li Hao, John Moore, Debraj Ray, Phil Reny, Tom Wiseman and seminar participants at various universities.

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

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

with his match. Each agent maximizes the discounted present value of expected monetary payoffs. Every relationship formed immediately generates positive payoffs for both parties. Each agent has a physical colour, either black or white. Before an agent decides whether or not to enter into a business relationship, he observes the physical colour of his potential partner and an additional piece of information about his match's past partners. 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, there is a chance that his social colour will switch to the physical or social colour of his partner.

The main result of this paper is that there exist equilibria which involve discrimination under certain conditions. In such environments we prove the existence of three types of discriminatory equilibria. One type involves segregation: members of each race discriminate against those of a different colour. In the other two types of equilibria, discrimination is one-sided: one race strongly discriminates against the other, while members of the persecuted race either use colour-blind strategies or only weakly discriminate.

The two prevalent economic theories of discrimination, taste-based and statistical, are based on attributes which directly affect payoffs either through preferences or production. Our theory is perhaps more related to a sizable literature in sociology which points to an alternative explanation of racial discrimination. This explanation is based on social norms rather than on payoff relevant characteristics. According to this theory, as in our model, members of a certain group practice prejudice and discrimination toward non-members because such behaviour is tolerated, and indeed expected, by other group members. Next, we discuss some evidence consistent with this theory.

A series of experiments conducted by Henri Tajfel suggests that discrimination and group identity might not be related to payoff relevant characteristics. Subjects developed a preference for their own group members even if the groups were artificially created (see Tajfel (1970), Tajfel and Turner (1979), Tajfel, Billig, Bundy, and Flament (1971), and Haslam (2004)). Teenage boys were assigned to one of two groups, [1] and had to choose between various monetary allocations among individuals. Most subjects exhibited a strong bias toward their own group. In particular, many subjects were willing to increase their group's allocation at the cost of reducing the total amount of money to be allocated. Some subjects were still willing to increase the difference between the total monetary payoffs allocated to the two groups even at the cost of reducing their group's allocation.

The famous Robbers Cave Experiment (see Sherif (1961)) also showed that prejudice and discrimination could be artificially created and manipulated by varying the social environment. In this experiment, teenage boys were randomly divided into two groups. In the first phase of the experiment, the groups were placed in competition with each other. After the competition started, subjects soon began to exhibit hostile behaviour towards members of the other group such as name-calling, singing derogatory songs, and refusing to eat together. In the second phase of

[1] In one of these experiments, for example, the subjects were asked to estimate the number of dots flashed on a screen. Then they were assigned one of two labels: "underestimators" and "overestimators."

the experiment, the two groups had to solve tasks cooperatively. This phase led to a dramatic improvement of the relationship between the two groups.

Minard (1952) finds that racial attitudes change according to the social environment. Black and white coal miners in West Virginia were monitored, and it was documented that they exhibited less hostility toward each other in their working environment than otherwise. For example, white workers were happy to sit next to black ones on miners' buses but they refused to sit next to the same workers on interstate buses. Pettigrew (1958) makes a similar observation about the difference between the attitudes of white southern military men while serving in the army and after they were discharged.

In our model, a white agent discriminates against black workers because he fears that if he did not, other white agents may refuse to hire him in the future. In other words, racist behaviour is sustained by the possibility of punishment by one's own group members. Indeed, peer pressure and the threat of social rejection are often mentioned as an explanation for bigotry (see Chapter 18 of Allport (1979)). Crossing group lines is often punished by stigmatization (see Austen-Smith and Fryer (2005) and the references therein), and sexual relationships or marriages outside of the community may lead to exclusion and even violence (Root (2001), Fryer (2007)).

In our theory, it is crucial that the social colour of a white agent can turn to black even if he employs another white agent with black social colour. In other words, one might become associated with the other race only indirectly. Such a technology makes it possible to punish not only those who do not discriminate but also those who do not punish non-discriminators. Are there social institutions with similar features? The Indian caste system, for example, prescribes several rules which prohibit certain kinds of relationships between members of different castes (see Pruthi (2004)). These rules are often enforced using the idea of *pollution*. Some castes are considered inherently polluted. A person who accepts a favour or food from a polluted person becomes polluted himself. That is, pollution is treated as something contagious which can only be cured by performing costly rituals.

The existing literature on taste-based discrimination is ample, see Becker (1971) and Schelling (1971). These approaches explain racial discrimination by assuming that individuals derive disutility from interacting with members of a different race. Such preferences may be the result of group selection; perhaps one group gains an advantage over other groups when its members cooperate only with each other and not with outsiders. Alternatively, a taste for discrimination might develop as an outcome of group formation processes. Similar people tend to have similar backgrounds, equipping them with similar tastes, values, and attitudes, and these shared qualities might facilitate collective decision making (Baccara and Yariv (2008), see also Alesina and Ferrara (2005)). Chapter 4 of Jones (1984) proposes a model involving a specific taste for conformism and which can be used to explain the dynamics of discrimination in the workplace.

A common critique of taste-based theories of discrimination is that employers who do not discriminate make larger profits than those who discriminate, hence the latter would not succeed

in competitive markets (see Becker (1971)). In our model, an employer who does not discriminate also has higher instantaneous profits. However, these short-term gains from a colour-blind hiring policy are offset by the boycott an employer will face from members of his own race in the future. That is, it is precisely the employer's profit-maximizing behaviour that leads to discrimination in equilibrium.

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

Arrow (1973) shows that discrimination can be a result of self-fulfilling expectations even if all agents are identical ex-ante. In his model, workers 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. Various extensions of the statistical discrimination theory include Coate and Loury (1993), Moro and Norman (2004), Rosén (1997), Mailath, Samuelson, and Shaked (2000) and Lang, Manove, and Dickens (2005).

In Gneezy, List, and Price (2012), field experiments across several market and agent-characteristics are used to identify the source of discriminatory behaviour. They find that when the object of discrimination is perceived to be controllable, discrimination is taste-based. If the object of discrimination is exogenous, discrimination is of statistical nature.

In our model, workers are identical both ex-ante and ex-post. Unlike the vertical discrimination caused by statistical discrimination, our setup might result in a mutual bias, with each race discriminating against the other. Such a phenomenon is inconsistent with statistical discrimination because the signal value of colour must be the same for any employer, regardless of his own colour.

Ours is not the first model in which discrimination arises without the presence of payoff-relevant differences between agents of different colours. Eeckhout (2006) considers a dynamic marriage market involving random matching of individuals. Once a marriage is formed, the two partners repeatedly play the Prisoner's Dilemma game. If either partner defects, both individuals return to the market and receive new matches. 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 immediately. The author shows that any colour-blind equilibrium is Pareto dominated by strategies in which the probability of defection depends on the colour of the partner.

The model presented by Mailath and Postlewaite (2006) involves a population of men and women who, each period, are matched and produce offspring. Agents differ in their non-storable endowments, and care about the consumption of their descendants. In addition, some agents have

a particular physical attribute, such as blue eyes, which is inherited by offspring. There exist equilibria in which the attribute has a value, that is, agents with the attribute are better off than agents without it. In this type of equilibrium, high-endowment agents without the attribute prefer to match with low-endowment agents with the attribute rather than with high-endowment agents without it. Such preferences arise from risk-aversion among agents; high-endowment individuals are willing to forgo 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. Because in our setup agents are risk-neutral, they have no incentive to transfer wealth across periods. However, while our concept of social colour is payoff-irrelevant, it acquires a value in equilibria, similar to the biological attribute in Mailath and Postlewaite (2006).

Peski and Szentes (2012) takes the analysis of our model one step further and characterizes the set of *stable* equilibria. They call an equilibrium stable if, after perturbing the equilibrium strategies slightly, myopic best-response dynamics imply convergence back to the equilibrium. The main result of Peski and Szentes (2012) is that each stable equilibrium involves discrimination under certain conditions. In particular, the colour-blind equilibrium is unstable. Under their assumed conditions there are three stable equilibria. One equilibrium involves segregation: members of each race discriminate against those of a different colour. In the other two equilibria, discrimination is one-sided: only one race discriminates against the other.

The social colour allocated to each agent in our model plays a role which is similar to the labels in Kandori (1992). Kandori considers a model in which members of two communities have repeated interactions. In every period, each member of a community is randomly matched with a member of the other community, and the pair plays a game. Players only observe the actions played in their past matches. However, each player is able to observe his partner's label, which contains some information about his past actions. An individual's label is updated each period, and is determined by his previous label, his partner's label, and the action he takes. Players might choose to condition their behaviour on labels, despite the fact that they are not directly payoff-relevant. The author proves a Folk Theorem for this setting. The label in our model, the social colour, does not only depend on actions but also on the physical characteristics of the players. Our paper also shows that acting on payoff-irrelevant information is a possibility. In addition, if one embraces the concept of myopic best-response stability, Peski and Szentes (2012) shows that in certain environments, stable equilibria *necessarily* involve discrimination.

2 The Model

Consider a population of agents, normalized to have unit mass. 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 ar-

rive independently across agents and time according to a Poison distribution with arrival rate δ. Agents with opportunities are matched into pairs instantaneously. Within a match, each agent is designated as either the *employer* or the *worker* with equal probability.[2] The two agents observe a match specific shock, s, which is exponentially distributed, that is, $G(s) = 1 - e^{-\lambda s}$. The employer then decides whether or not to employ the worker. If he does employ 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. Every 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 is his *social colour*. The physical colour is either black (b) or white (w), and is immutable. A fraction μ_w of the population is white, while the remaining fraction $\mu_b\,(= 1 - \mu_w)$ is black. An agent's social colour is also either black or white, and evolves as follows. The social colour of a worker remains unaffected by his match.[4] If an employer employs a worker with type (c_1, c_2), the employer's social colour remains unchanged 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 the worker, his social colour remains unchanged 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, in the past, he hired white workers or workers with white social colour.

Agents' types are payoff irrelevant in the following sense. An agent's payoff depends only on the history of shock realizations and employment decisions in his past relationships, but not on his type, nor on the types of agents with whom he interacts. If there were no types, this model would have a unique equilibrium in which employers always choose to employ whichever workers they are matched with. In fact, this is true even if agents have physical colours but no social colours; an employer receives a positive payoff if he employs the worker and, in the absence of social colour, such a decision cannot affect his future employment.

In this model, only employers make decisions. An employer's strategy is a mapping from his private history, his type, the type of the worker, and the realization of the shock into an employment decision. In what follows, we restrict our attention to *steady state equilibria*. That is, we characterize equilibria in which (i) the agents' strategies depend neither on time nor on his private history, and (ii) the distribution of types is constant over time.

[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] Since s is always positive, the total surplus generated in a relationship, $s + M$, is strictly positive.

[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.

3 Best Responses

This section characterizes the employers' best-response decisions. An employer's optimal hiring decision is a complicated object even in a stationary environment because it might depend on his type, the type of the worker and the realization of the shock. Nevertheless, we are able to reduce the complexity of the employer's problem appreciably. First, note that the optimal hiring decision can always be characterized by cutoffs; 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. These cutoffs can depend on the types of both the employer and the worker, so there might be sixteen of them. Second, we will show 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. Finally, we will prove that the various cutoffs of a black (white) employer are linearly dependent on one another, with coefficients determined by the parameters of our model. This implies that any one of the cutoffs completely determines the values that the other three cutoffs will take. As a consequence, the best-response decision of a black (white) agent can always be represented as a one-dimensional variable. This result is significant in the sense that finding a stationary equilibrium is now reduced to a two-dimensional problem; we need to characterize one cutoff for each race.

In the remainder of this section, we characterize the equilibrium values in terms of the two relevant 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

In what follows, we derive the initial best-response cutoffs of an agent against an arbitrary population strategy. To this end, we fix a population strategy and a distribution of types at time zero. Let V_{c_1,c_2} denote the value of an agent with type (c_1, c_2) $\left(\in \{b, w\}^2\right)$ at time zero, before he knows whether a production opportunity has arrived. That is, V_{c_1,c_2} is the maximum discounted present value of the payoffs that a type-(c_1, c_2) agent can achieve given the strategy and type-distribution of the others. This value depends only on type and not on the identity of the agent, because two agents with the same type face the same environment.

For example, the optimal cutoff for a white employer with social colour c who presently faces a worker with type (b, w) is computed as follows. Suppose that the value of the shock is s. If he employs the worker, 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 discounted present value is equal to

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

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 $\max\{0,\gamma\alpha\,(V_{w,w}-V_{w,b})\}$.

Note 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 $(1-\gamma)\,V_{w,c}$, which cancels out in the computation of the cutoff. In fact, an employer's social colour only affects his payoff in the event that his social colour remains unchanged, and this event is independent of his decision. Therefore, while the best-response cutoff of an agent may depend on his physical colour, it cannot depend 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:*

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

An employer with physical colour c who is considering hiring a worker will be concerned about the effect it will have on his 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* the agent has towards his own physical colour.[5] The above lemma implies that the best-response cutoffs are proportional to this bias, up to the requirement that the cutoffs be non-negative. The coefficients of the bias corresponding to 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 note that

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

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

Since agents of the same type have identical values, Lemma 1 implies that any stationary equilibrium is symmetric. That is, employers with the same physical colour use the same strategies. Also note 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 used against different types of workers are defined by (3).

In order to show that a cutoff profile (x_*^c, x_*^{-c}) constitutes an equilibrium, we need to prove that there is a distribution of social colours such that (i) x_*^c is a best response cutoff of an agent with colour c against the population strategy profile (x_*^c, x_*^{-c}) for $c \in \{b, w\}$ given the distribution of types, and (ii) the distribution of social colours does not change over time. The next section characterizes the best response correspondence, and in particular, shows that the best response of an agent only depends on the cutoffs of others but not on the distribution of social colours.

3.2 The Best-Response Curves

Our next goal 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 $(-c)$ always uses cutoff x^c (x^{-c}). This notation presumes that the best responses, $b^c\left(x^c, x^{-c}\right)$, does not depend on the distribution of social colours. The next lemma implies that this is indeed true.

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

$$b^c\left(x^c, x^{-c}\right) = K \max\left\{0, \mu_c G\left(\left(1-\alpha\right)x^c\right) + \mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)\right\}, \tag{4}$$

where $K = M\delta\gamma/\left(2r + \gamma\delta\right)$.

The proof of this lemma is relegated to the appendix and requires the computation of the bias, $V_{c,c} - V_{c,-c}$, for $c \in \{b, w\}$ given $\left(x^b, x^w\right)$. These objects then identify the best-response cutoffs by Lemma 1. Before we establish some formal results concerning the function b^c, it is worth discussing a few of its attributes.

The function b^c is decreasing in the discount rate r. An agent might elect to not hire a worker, sacrificing an instantaneous payoff, in order to affect his social colour, which in turn impacts his future payoffs. The more patient an agent is, the larger the payoff he is willing to forgo at present in exchange for a higher continuation payoff.

The social colour of an agent only matters in the event that he is a worker. If the wage of a worker, M, increases, it becomes more beneficial to have the social colour desired by employers. So, the higher the wage, the larger are the cutoffs used by an agent in order to increase the probability of his future employment. Therefore, the function b^c is increasing in M.

An agent does not care directly about the others' cutoffs, he only cares about the probability of being hired. Recall that $G\left(s\right) = 1 - e^{-\lambda s}$. It is easy to verify that b^c is increasing in λ. The larger the parameter λ, the more likely it is that the realization of the shock is small. Therefore, given

(x^c, x^{-c}), a larger λ increases the likelihood that an agent will actually be affected by the cutoffs of the others. Hence, if λ is large, the agent is willing to use large cutoffs in order to increase the probability of future employment.

A notable feature of the best response function, b^c, is that it does not depend on the distribution of social colours. Recall from Lemma 1 that the best-response cutoff of an agent with physical colour c only depends on the bias $V_{c,c} - V_{c,-c}$. Note that b^c defines the best-response cutoff of an agent with colour c if each employer with physical colour c $(-c)$ always uses the same cutoffs which satisfy (3). In fact, as we shall now explain, the bias does not depend on the population distribution of social colours as long as the agents' cutoffs satisfy (3). An agent's value consists of the discounted sum of instantaneous payoffs he receives as a worker and as an employer in the future. Since the cutoffs of other employers satisfy (3), they do not depend on their social colours and hence, the expected instantaneous payoff of a worker is also independent of the distribution of the social colours. On the other hand, the same agent's payoff as an employer does depend on the distribution of social colours because his decision to hire depends on the worker's type. However, since his cutoffs do not depend on his own social colour (see Lemma 1), the terms in $V_{c,c}$ and $V_{c,-c}$ which depend on the type distribution are the same and they cancel out in the computation of $V_{c,c} - V_{c,-c}$. For this reason, we do not need to compute the stationary distribution of types in order to analyse the equilibrium behaviour. The definition of stationary equilibrium requires a stationary type distribution and the existence of such a distribution given the stationary strategies is obvious.

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

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

Figure 1 plots $b^b(., 0)$ and $b^b(., x^w)$ $(x^w > 0)$ for the case when λ is large. For $x^w > 0$, $b^b(., x^w)$ is a downwards shift of $b^b(., 0)$, or zero if the shifted curve becomes negative. The function $b^b(., x^w)$ is locally concave in x^b whenever it is positive (part (i) of Lemma 3). Part (ii) of Lemma 3 states 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 is always better off employing 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 whether white agents do not discriminate $(x^w = 0)$ or whether they discriminate fully $(x^w = \infty)$. The reason for this is that a black agent is always employed by white agents if $x^w = 0$ and is never

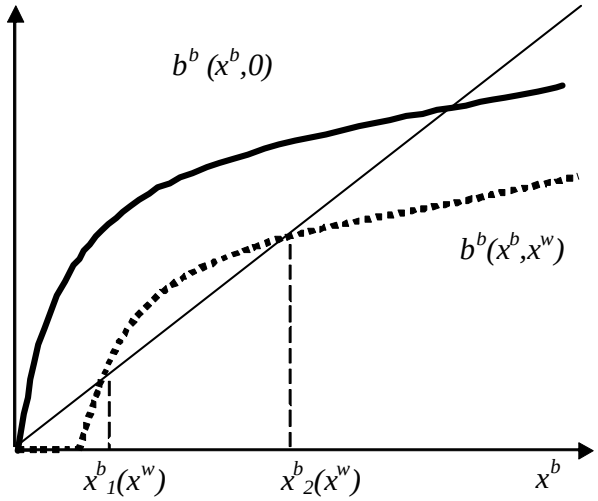

Figure 1: Best Responses for Large λ

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 $b^b\left(x^b, x^w\right)$ 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 in Figure 1 the curve $b^b(.,0)$ intersects the 45-degree line twice. Part (ii) of the previous lemma only implies that the function $b^c(x^c,0)$ intersects the 45-degree line at $x^c = 0$. Whether or not there is another intersection carries great importance in characterizing the set of equilibria. It turns out that the existence of a strictly positive intersection largely depends on the relative fraction of payoffs due to being a worker and being an employer. Ceteris paribus, as λ increases, labour income increases relative to the employer's income. Similarly, if the wage M increases, it becomes more important to be employed relative to being an employer. In what follows, we investigate the set of equilibria as a function of the parameter λ. Alternatively, we could have presented our results in terms of the size of M.

Recall that K denotes the multiplicative coefficient of b^c in (4).

Lemma 4 *Let* $\lambda_0 = 1/\left(K\left(1-\alpha\right)\mu_c\right)$. *Then,*

(i) if $\lambda > \lambda_0$, *then there exists a unique* $x^c > 0$ *such that* $b^c\left(x^c,0\right) = x^c$, *and*

(ii) if $\lambda \leq \lambda_0$, *then* $b^c\left(x^c,0\right) < x^c$ *for all* $x^c > 0$.

The function $b^b(.,0)$ is strictly concave and zero at the origin. In addition, this slope converges to zero as x^b goes to infinity. Therefore, the function $b^b(.,0)$ intersects the 45-degree line at a strictly positive value if and only if its slope at zero is larger than one. As we explained before, $b^b\left(x^b, x^w\right)$ is increasing in λ (see the discussion after Lemma 2). Since $b^b(0,0) = 0$, it follows that the slope of $b^b(.,0)$ is large if and only if λ is large. At the critical value $\lambda = \lambda_0$, the slope of the best-response function, $b^b(.,0)$, is exactly one.

4 Equilibria

This section accomplishes two goals. First, we give an exact characterization of those environments where the colour-blind equilibrium is not the unique equilibrium. To be more specific, we show that there exist equilibria in which some agents discriminate if and only if $\lambda > 1/\left(K(1-\alpha)\mu_c\right)$ for some $c \in \{b, w\}$. Second, we give a graphical representation of the set of equilibria and prove the existence of various types of equilibria if λ is large. In particular, we show that there is an equilibrium in which the two races segregate, that is, both x^c and x^{-c} are large.

The existence of a discriminatory equilibrium hinges on the size of the parameter λ. Suppose for a moment that white agents are non-strategic and that their cutoff is zero, and consider a version of our model in which the game is played only by black agents. As previously explained, the larger the parameter λ, the more likely it is that the realization of the shock is small. Therefore, given the cutoff used by other black agents, a larger λ increases the likelihood that a black agent with a white social colour will not be employed by other black agents in the future. Hence, even if the black agents' cutoff is small, a large λ induces a black agent to respond with large cutoff in order to maintain a black social colour. In other words, if the black agents' cutoff is small but λ is large, a black agent's best response is larger than the others' cutoff. This is why in Figure 1 the thick curve is above the 45-degree line near zero. On the other hand, since the continuation payoff is bounded, the best-response cutoff is also bounded from above. So, if the others' cutoff is very large, the best-response cutoff of an agent is smaller than that of the others. This explains why, in Figure 1 , for large values in the domain the thick curve is below the 45-degree line. By continuity, there exists a cutoff to which the best response is the cutoff itself, that is, a cutoff which defines an equilibrium. This cutoff corresponds to the intersection of the thick curve and the 45-degree line on Figure 1. Now, let us allow white agents to be strategic. As we previously pointed out (see part (ii) of Lemma 3), the best response of a white agent is zero as long as the other white agents' cutoff is zero. Therefore, the equilibrium we have just identified remains an equilibrium even if white agents choose their cutoffs strategically.

The argument of the previous paragraph does not depend on our specific assumption on the distribution of the shock. In general, a discriminatory equilibrium exists if the slope of the best-response curve is larger than one at zero, which in turn, depends on the density of the shock distribution at zero. It is not hard to prove that if distribution of the shock has a positive density,

g, on $\mathbb{R}_+$, a discriminatory equilibrium exists whenever $g(0)$ is sufficiently large.

Proposition 1 *The colour-blind cutoff profile,* $(0,0)$*, is an equilibrium. In addition,*

(i) if $\lambda \leq 1/(K(1-\alpha)\mu_c)$ *for both* $c \in \{b,w\}$*, the profile* $(0,0)$ *is the unique equilibrium, and*

(ii) if $\lambda > 1/(K(1-\alpha)\mu_c)$ *for either* $c = b$ *or* $c = w$*, there exists an equilibrium* (x_*^c, x_*^{-c}) *such that* $x_*^c > 0$*.*

Before we prove this proposition, we restate the definition of equilibrium in terms of the best-response curves as follows. The cutoff profile (x_*^c, x_*^{-c}) is an equilibrium if and only if

$$\left(x_*^c, x_*^{-c}\right) = \left(b^c\left(x_*^c, x_*^{-c}\right), b^{-c}\left(x_*^{-c}, x_*^c\right)\right). \tag{5}$$

Proof. Recall that part (ii) of Lemma 3 says that $b^c(0, x^{-c}) = 0$ for all x^{-c} and $c \in \{b,w\}$. In particular, $b^c(0,0) = 0$ for $c \in \{b,w\}$. Hence, $(0,0)$ satisfies (5).

In order to prove part (i) we have to show that if $\lambda \leq 1/(K(1-\alpha)\mu_c)$ for $c \in \{b,w\}$ then the only equilibrium is $(0,0)$. Suppose that (x_*^c, x_*^{-c}) is an equilibrium. Then equation (5) implies that $b^c(x_*^c, x_*^{-c}) = x_*^c$ for $c \in \{b,w\}$. Since $\lambda \leq 1/(K(1-\alpha)\mu_c)$ it follows from part (iii) of Lemma 3 and part (ii) of Lemma 4 that $x_*^c = 0$ for $c \in \{b,w\}$.

We turn our attention to part (ii). If $\lambda > 1/(K(1-\alpha)\mu_c)$ then there exists a unique $x^c > 0$ such that $b^c(x^c, 0) = x^c$ by part (i) of Lemma 4. In addition, $b^{-c}(x^{-c}, 0) = 0$ by part (ii) of Lemma 3. Therefore, $(x^c, 0)$ satisfies equation (5). ■

Part (ii) of Proposition 1 provides little information about the set of equilibria which involve discrimination. Next, we provide a graphical characterization of the equilibria for the case of a large λ.

Note that if (x_*^c, x_*^{-c}) is an equilibrium, then by (5), $x_*^c = b^c(x_*^c, x_*^{-c})$ for $c \in \{b,w\}$. This means that the function $b^c(., x_*^{-c})$ intersects the 45-degree line at x_*^c. As previously indicated (see part (ii) of Lemma 3), these curves intersect at zero. Next, we investigate intersections which are strictly positive.

We will show that for a generic x^{-c}, there are either two positive intersections of $b^c(., x^{-c})$ and the 45-degree line, or there are none.[6] Figure 1 illustrates a situation in which there are two intersections for $c = b$. (These intersections are denoted by $x_1^b(x^w)$ and $x_2^b(x^w)$.) For each x^{-c}, let $x_1^c(x^{-c})$ and $x_2^c(x^{-c})$ denote the smaller and larger positive intersections respectively, if they exist. We will argue that, depending on the parameter values, there are two different cases which can arise. Case 1: there are two intersections for each x^{-c} and hence, x_1^c and x_2^c are defined everywhere. Case 2: there are two intersections if $x^{-c} \notin \left(\underline{x}^{-c}, \overline{x}^{-c}\right)$ and there is no intersection if $x^{-c} \in \left(\underline{x}^{-c}, \overline{x}^{-c}\right)$. In this case, the curves x_1^c and x_2^c are only defined on $\mathbb{R}_+ \setminus \left[\underline{x}^{-c}, \overline{x}^{-c}\right]$. The next figure depicts x_1^c and x_2^c for both cases.

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

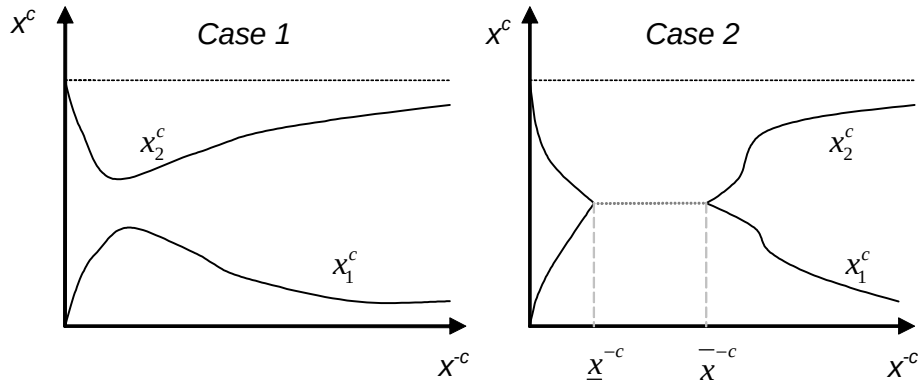

Figure 2: Positive Intersections

Next, we explain how the curves are drawn on Figure 2. Recall that the function $b^c(.,x^{-c})$ is essentially a downward shift of $b^c(.,0)$ (see Figure 1). The size of this shift determines the number of positive intersections. It is easy to show that this size is a non-monotonic function of x^{-c}. If x^{-c} is small, an increase in x^{-c} shifts the curve $b^c(.,x^{-c})$ even further down. Above a certain value of x^{-c}, however, a further increase in x^{-c} shifts the curve $b^c(.,x^{-c})$ upwards. In fact, as x^{-c} goes to infinity, $b^c(.,x^{-c})$ converges back to $b^c(.,0)$ (see part (iii) of Lemma 3). Recall that if λ is large, the first derivative of $b^c(.,0)$ is larger than one (see Lemma 4). Hence, $b^c(.,x^{-c})$ and the 45-degree line have two intersections if the downward shift is small, and none if the shift is large. In the latter case, $b^c(.,x^{-c})$ is pushed below the 45-degree line. Case 1 corresponds to parameters where the curve $b^c(.,x^{-c})$ intersects the 45-degree line even when it is shifted furthest down. In Case 2, there is an interval such that, if x^{-c} lies in this interval, the curve $b^c(.,x^{-c})$ is pushed below the 45-degree line. If x^{-c} is outside of this interval, there are two positive intersections.

In both cases the curve x_1^c first increases then decreases, because, the larger the downward shift, the higher the first point of positive intersection will be. Similarly, the curve x_2^c decreases first, then increases because the location of the second positive intersection decreases as the size of the shift increases. In the panel corresponding to Case 2, the values of x_1^c and x_2^c are equal at $\underline{x}^{-c}$ and $\overline{x}^{-c}$; both $\underline{x}^{-c}$ and $\overline{x}^{-c}$ induce the same shift, that is, $b^c(.,\underline{x}^{-c}) = b^c\left(.,\overline{x}^{-c}\right)$. In addition, the shifted best-response curve is exactly tangent to the 45-degree line, hence, the two intersections collapse into one. These results are stated and proved in the working paper version of our paper (see Lemma 5 in Peski and Szentes (2012)).

Since $b^c(.,x^{-c})$ intersects the 45-degree line at zero for all x^{-c} (see part (ii) of Lemma 3), the curve $x_0^c(x^{-c}) \equiv 0$ 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 $\{x_i^w\}_{i=0}^2$. Formally, (x_*^c, x_*^{-c}) is an equilibrium cutoff profile

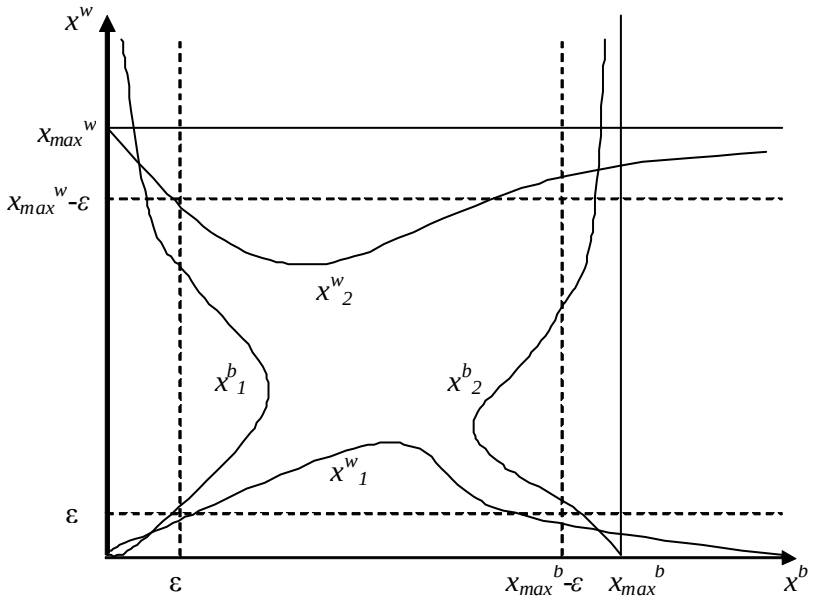

Figure 3: Equilibria

if and only if there exist $i, j \in \{0, 1, 2\}$ such that

$$x_*^c = x_i^c\left(x_*^{-c}\right) \text{ and } x_*^{-c} = x_j^{-c}\left(x_*^c\right). \tag{6}$$

Therefore, in order to find equilibria geometrically, we need to add the curves $\left\{x_i^{-c}\right\}_{i=0}^2$ to Figure 2 and find every intersection. We did exactly this in Figure 3, in an environment where both colours satisfy Case 1 in Figure 2.

Intuitively, the curve x_2^c corresponds to strong discrimination of agents with colour c,, x_1^c corresponds to weak discrimination, and x_0^c implies no discrimination. In Figure 3 any combination of these behaviours is present in an equilibrium. Indeed, each curve x_i^c $(i = 0, 1, 2)$ intersects with each curve x_j^{-c} $(j = 0, 1, 2)$. If one of the colours satisfied Case 2 in Figure 2, some of these intersections might not exist. In general, we claim neither the existence, nor the uniqueness of these intersections. In the proof of Proposition 1, however, we showed that the intersection of x_0^c and x_2^{-c} exists and is unique ($c \in \{b, w\}$) if $\lambda > 1/\left(K\left(1-\alpha\right)\mu_c\right)$. The unique intersection of x_0^c and x_0^{-c}, $(0, 0)$, corresponds to the colour-blind equilibrium. Next, we also show that if λ is large then the intersection of x_2^c and x_2^{-c} exists. That is, there always exists an equilibrium where agents segregate and discriminate strongly.

Note that by (4), 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

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

Proposition 2 *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 $x^c_* \in [0,\varepsilon)$ or $x^c_* \in (x^c_{\max}-\varepsilon, x^c_{\max})$. In addition, there exists equilibria with each of the following properties:*

(i) $x^c_* \in [0,\varepsilon)$ and $x^{-c}_* \in [0,\varepsilon)$,

(ii) $x^c_* \in [0,\varepsilon)$ and $x^{-c}_* \in (x^{-c}_{\max}-\varepsilon, x^{-c}_{\max})$,

(iii) $x^{-c}_* \in [0,\varepsilon)$ and $x^c_* \in (x^c_{\max}-\varepsilon, x^c_{\max})$, and

(iv) $x^c_* \in (x^c_{\max}-\varepsilon, x^c_{\max})$ and $x^{-c}_* \in (x^{-c}_{\max}-\varepsilon, x^{-c}_{\max})$.

This proposition states that if λ is large enough, then, in every equilibrium, an agent either has a small cutoff ($x^c_* < \varepsilon$), or a very large cutoff ($x^c_{\max}-\varepsilon < x^c_*$). In addition, any combination of these strategies can arise in an equilibrium. That is, there are equilibria where both races use small cutoffs, there are equilibria where one race discriminates strongly against the other while the other race hardly discriminates, and there equilibria where both races discriminate strongly.

In the proof of this proposition, we show that an agent who uses a cutoff $x^c_* \in (x^c_{\max}-\varepsilon, x^c_{\max})$ employs a worker with physical or social colour $-c$ with small probability. In fact, we prove that as λ goes to infinity the probability of employment and the expected surplus generated by such matches goes to zero. Similarly, we show that if an agent uses a cutoff $x^c_* \in (0,\varepsilon)$, then he employs a worker with high probability irrespective of the worker's type. As λ goes to infinity, the probability of employment induced by a weakly discriminating cutoff goes to one.

The critical value of λ above which the various equilibria described in parts (i)-(iv) of Proposition 2 exist is strictly larger than the threshold value of λ from part (ii) of Proposition 1. To see this, let us explain that λ needs to be larger to sustain a two-sided discriminatory equilibrium (part (iv) of Proposition 2) than $1/\left(K\left(1-\alpha\right)\mu_c\right)$, which guarantees the existence of a discriminatory equilibrium where only colour-c agents discriminate (part (ii) of Proposition 1). Consider a cutoff profile (x^c_*, x^{-c}_*) where $x^c_*, x^{-c}_* > 0$. By (5), in order for (x^c_*, x^{-c}_*) to be an equilibrium, the equation $b^c\left(x^c_*, x^{-c}_*\right) = x^c_*$ must hold, that is, the curve $b^c\left(., x^{-c}_*\right)$ intersects the 45-degree line at x^c_*. As we explained earlier (see part (iii) of Lemma 3), the curve $b^c\left(., x^{-c}_*\right)$ is essentially a downward shift of $b^c\left(., 0\right)$. This is because as x^{-c} jumps from zero to x^{-c}_*, a colour-c worker benefits from having a social colour $-c$ whenever he meets an employer with physical colour $-c$, which, in turn, induces him to decrease his cutoff. Recall that if $\lambda = 1/\left(K\left(1-\alpha\right)\mu_c\right)$, the slope of $b^c\left(., 0\right)$ is exactly one at zero. Therefore, if λ is just above $1/\left(K\left(1-\alpha\right)\mu_c\right)$ then $b^c\left(., 0\right)$ intersects the 45-degree line but the shifted curve $b^c\left(., x^{-c}_*\right)$ will not. So, in order to have a two-sided discriminatory equilibrium we need a λ large enough so that even the shifted curves, $b^c\left(., x^{-c}_*\right)$ and $b^{-c}\left(., x^c_*\right)$, intersect the 45-degree line.

5 Discussion

Next, we discuss some assumptions and extensions of our model. In this paper, our goal was to present a simple model which demonstrates that discrimination can arise purely because agents observe information about others' past actions. We do not claim that equilibrium discrimination is robust to all possible modifications of our model. However, some of the assumptions we make are only necessary in order to provide a graphical representation of the equilibria.

Distribution of shocks.– We have assumed that the match-specific shock that determines the surplus of a partnership is exponentially distributed. Whether or not the colour-blind equilibrium is unique depends only on the slopes of the best-response functions at $(0, 0)$. These slopes are increasing functions of the density of the shock distribution at zero. Whenever this density is large there exists a discriminatory equilibrium.

Note 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 negative shocks were allowed, the uniqueness of the colour-blind equilibrium would only depend on the slope of the best-response curve at the origin. However, in this case, there might be equilibria different from ours. In particular, it is possible that white employers would prefer to hire black workers and vice versa, that is, it could be more valuable to have a social colour which is different from one's physical colour.

Social colour.– If an employer chooses not to hire, then, if his social colour changes it will change to his own physical colour. The motivation for this assumption is that if a white agent refuses to hire a black employee despite the positive surplus, he will be viewed as 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 social colour is a binary signal and its evolution is only determined by the physical colours of the worker and the employer; this is in the spirit of Kandori (1992). One might choose to model the information an employer observes about a worker in a more complicated way. For example, an employer might draw a random sample of the physical colours of the agents in the worker's partnership history. Then, the type of the worker would consist of both his physical colour and his full history. Such modelling would lead to a complex type space but does not alter our main results.

It is easy to construct social colours different from ours which do not lead to discrimination. For example, if the colour is not informative about past decisions then the colour blind equilibrium is unique. The characterization of those processes which might lead to discrimination is beyond the scope of this paper.

More attributes and social colours.– In reality, individuals have more than one physical attribute. It is also possible that an individual is subject to several labels that depend 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 and λ is large, discriminatory equilibria will exist.

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 employed too infrequently. It would be interesting to allow wages to be endogenous and analyse wage differentials due to racial discrimination. Unfortunately, it is not entirely clear how endogenous wages would affect our main results. Difficulties arise from the fact that if a black worker is willing to take a paycut in order to be employed by a white employer, more white employers will employ black workers. This would increase the number of white agents with black social colour, which in turn, would make it less costly for a white agent to have a black social colour. Therefore, it would be less likely for discrimination to arise in equilibrium. A potential solution to this problem would be to allow social colour to change as a function of the wage offered to a worker, for example, a lower wage for a black worker could lead to an increased likelihood that the employer's social colour becomes white.

We are currently developing 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 Conclusion

This paper puts forward a new theory of racial discrimination. Individuals discriminate because they do not want to be associated with the other race. Although the information about others' association is not payoff-relevant, it plays a major role in determining the behaviour of economic agents.

Our model does not attempt to explain why agents might use skin colour as a basis for discrimination as opposed to other observable physical attributes. People differ in height, weight, eye-colour, and along many other dimensions. One potential explanation might take into account the fact that members of a family or a community are more likely to have the same skin colour than the same height or weight. Discrimination against short individuals might be difficult to sustain if many relatives of tall people are short. Recall that in our model, a white agent discriminates against those who associate themselves with blacks because he is afraid of those whites who associate more closely with whites. Since individuals must necessarily associate with short and tall

individuals, these attributes cannot be used to sustain discrimination. Another reason for using skin colour is because it is more easily observed that other attributes such as eye-colour.

Throughout the paper, we have assumed that the surplus generated by a partnership is exogenously divided between the worker and the employer. We have excluded the possibility that discrimination results in a wage differential. Perhaps the most important elaboration of our model would be to allow wages and profits to be determined endogenously.

We have not yet discussed policy in this paper. Recall that a white employer discriminates against black workers because he is afraid of being turned down by white employers with white social colour in the future. Hence, a policy intervention which would reduce the incentive to discriminate might involve increasing the fraction of the population whose social colours are different from their own physical colours. It is clear that subsidizing employers who hire workers of a different physical colour would increase the fraction of the population whose physical and social colour don't match. This would of course result in a lower proportion of individuals with the same physical and social colour, and reduce the incentive to discriminate. Such subsidies must be paid from taxes, which might alter the incentives to produce. Therefore, in order to discuss policy in a meaningful way, one must model production and the worker's incentives carefully.

7 Appendix

7.1 Proof of the Lemmas

Proof of Lemma 2. For each (x^c, x^{-c}) we shall compute the bias $V_{c,c} - V_{c,-c}$ for $c \in \{b, w\}$.[7] 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 agent's value function if he is a worker or employer respectively, where $(c_1, c_2) \in \{b, w\}^2$ is his type. 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:

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

To see this, notice that the probability a particular agent does not receive an opportunity in time dt is $1 - \delta dt$, and hence his value remains V_{c_1,c_2}. This is discounted at the rate r. Otherwise the agent receives an opportunity, and is equally likely to become an employer or a worker. After dividing through by dt and taking the limit as dt goes to zero, we obtain

$$V_{c_1,c_2} = \frac{\delta}{\delta + r} \left(\frac{1}{2}\Pi^l_{c_1,c_2} + \frac{1}{2}\Pi^e_{c_1,c_2} \right). \tag{7}$$

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 is also 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

[7] The obvious dependence of the values of the agents on (x^b, x^w) is supressed from the notation V_{c_1,c_2} for simplicity.

$\mu_{-c}\left(1-G\left(x^{-c}\right)\right)$. Finally, an employed worker's value changes to $V_{c,c}$, and he also receives M whenever he is employed, therefore,

$$\Pi_{c,c}^{l}=M\left(\mu_{c}+\mu_{-c}\left(1-G\left(x^{-c}\right)\right)\right)+V_{c,c}. \tag{8}$$

Similarly,

$$\Pi_{c,-c}^{l}=M\left(\mu_{c}\left(1-G\left(\left(1-\alpha\right)x^{c}\right)\right)+\mu_{-c}\left(1-G\left(\alpha x^{-c}\right)\right)\right)+V_{c,-c}. \tag{9}$$

Because the optimal hiring decision of employer with physical colour c does not depend on his social colour, the difference between value functions of employers with the same physical colour c is equal to the difference between the value functions of an unmatched agent multiplied by the probability that the social colour of the employer does not change:

$$\Pi_{c,c}^{e}-\Pi_{c,-c}^{e}=\left(1-\gamma\right)\left(V_{c,c}-V_{c,-c}\right). \tag{10}$$

Using (7), (8), (9), and (10) we can express $V_{c,c}-V_{c,-c}$ as follows:

$$\begin{aligned}V_{c,c}-V_{c,-c} &= \frac{\delta}{\delta+r}\left[\frac{1}{2}\left(\Pi_{c,c}^{l}-\Pi_{c,-c}^{l}\right)+\frac{1}{2}\left(\Pi_{c,c}^{e}-\Pi_{c,-c}^{e}\right)\right]\\ &= \frac{\delta}{\delta+r}\frac{1}{2}M\left[\mu_{c}G\left(\left(1-\alpha\right)x^{c}\right)+\mu_{-c}\left(G\left(\alpha x^{-c}\right)-G\left(x^{-c}\right)\right)\right]\\ &\quad+\frac{2-\gamma}{2}\frac{\delta}{\delta+r}\left[V_{c,c}-V_{c,-c}\right]\end{aligned}$$

That is,

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

Recall from Lemma 1 that the best-response cutoff of an employer with physical colour c against a worker with type $(-c,-c)$ is $\gamma\left(V_{c,c}-V_{c,-c}\right)$. Then the previous displayed equality implies that the best-response cutoff is

$$\widetilde{b}^{c}\left(x^{c},x^{-c}\right)=K\left[\mu_{c}G\left(\left(1-\alpha\right)x^{c}\right)+\mu_{-c}\left(G\left(\alpha x^{-c}\right)-G\left(x^{-c}\right)\right)\right], \tag{11}$$

where K denotes $M\delta\gamma/\left(2r+\gamma\delta\right)$. 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 (4). ■

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 and strictly increasing in x^{c}. By (11)

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

where $g\left(x\right)=\lambda e^{-\lambda x}$ for all $x\geq0$. This partial derivative is positive and decreasing.

(ii) By (11),

$$\widetilde{b}^{c}\left(0,x^{-c}\right)=K\left[\mu_{-c}\left(G\left(\alpha x^{-c}\right)-G\left(x^{-c}\right)\right)\right]\leq0,$$

because $G\left(\alpha\overline{x}^{-c}\right)-G\left(\overline{x}^{-c}\right)\leq0$. Hence, (4) and (11) imply $b^{c}\left(0,x^{-c}\right)=0$.

(iii) Notice that $\lim_{x^{-c}\to\infty} G(\alpha x^{-c}) - G(x^{-c}) = 0$. Therefore, by (4) and (11),

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

Finally, the inequality $b^c(x^c, 0) \geq b^c\left(x^c, \overline{x}^{-c}\right)$ follows from $G\left(\alpha\overline{x}^{-c}\right) - G\left(\overline{x}^{-c}\right) \leq 0$. ■

Proof of Lemma 4. Since $\widetilde{b}^c(x^c, 0) \geq 0$ by (11), (4) implies $\widetilde{b}^c(x^c, 0) = b^c(x^c, 0)$. Therefore, by the proof of part (i) of Lemma 3

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

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

(i) If $\lambda > \lambda_0$ then $K\mu_c(1-\alpha)\lambda > 1$. This means that $\partial b^c(x^c, 0)/\partial x^c\rfloor_{x^c=0} > 1$ and therefore, $b^c(x^c, 0) > x^c$ if x^c is close to zero. Since the curve $b^c(x^c, 0)$ 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 $b^c(x^c, 0) = x^c$.

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

7.2 Proof of Proposition 2

Before we proceed with the proof of Proposition 2 we prove a few Lemmas about the equilibrium cutoffs. For convenience we introduce a few new notations. We shall denote $\min\{\mu_b, \mu_w\}$ by $\mu_{\min}$. In addition, we define two constants

$$\psi_0 = \frac{1}{\frac{1}{4}K\alpha(1-\alpha)\mu_{\min}},\ \psi_1 = K\mu_{\min}\frac{1}{2}\left(1-2^{-\alpha}\right)\left(1-2^{-(1-\alpha)}\right).$$

In the proofs of the lemmas we often use the following inequality

$$1 - e^{-\xi} \geq \frac{1}{2}\xi \text{ for all } \xi \leq \log 2. \tag{12}$$

In what follows $\left(x^b, x^w\right)$ denote the equilibrium cutoffs, i.e.,

$$x^c = b^c\left(x^c, x^{-c}\right) \text{ for each } c. \tag{13}$$

Lemma 5 *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, because $x^c = \widetilde{b}^c(x^c, x^{-c})$ for each c (see (11)),

$$\begin{aligned} x^b + x^w &= \sum\nolimits_{c\in\{b.w\}} K\left[\mu_c G\left((1-\alpha)x^c\right) + \mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)\right] \\ &= \sum\nolimits_{c\in\{b.w\}} K\mu_c\left[G\left((1-\alpha)x^c\right) + G\left(\alpha x^c\right) - G\left(x^c\right)\right] \\ &= \sum\nolimits_{c\in\{b,w\}} K\mu_c\left(1-e^{-\alpha\lambda x^c}\right)\left(1-e^{-(1-\alpha)\lambda x^c}\right). \end{aligned} \tag{14}$$

The first equality above follows from rearranging the terms corresponding to the same colour. The second one follows 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 $\max\left\{x^b, x^w\right\} \geq (\log 2)/\lambda$, then it follows from the previous equality that

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

Since $\max\left\{x^b, x^w\right\} \geq \frac{1}{2}\left(x^b + x^w\right)$, 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 inequality (12),

$$1 - e^{-\alpha\lambda x^c} \geq \frac{1}{2}\alpha\lambda x^c \text{ and } 1 - e^{-(1-\alpha)\lambda x^c} \geq \frac{1}{2}(1-\alpha)\lambda x^c \tag{15}$$

for each $c \in \{b, w\}$. Equations (14) and inequalities (15) imply that

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

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

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 (13),

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

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

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

If $x^w \leq (\log 2)/\lambda$ then, by inequality (12),

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

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

Lemma 6 *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 $x^c \geq \psi_1$. Suppose that $x^{-c} > 0$. Then

$$\begin{aligned} x^{-c} &= K\mu_{-c}G\left((1-\alpha)x^{-c}\right) + K\mu_c\left(G(\alpha x^c) - G\left(x^c\right)\right) \\ &\geq K\mu_{-c}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) - K\mu_c e^{-\lambda\alpha\psi_1}, \end{aligned} \tag{16}$$

where the equality is just (13) and the inequality follows from $x^c \geq \psi_1$. We consider two cases.

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

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

where the last equality follows from $\psi_1 \leq K\mu_{-c}\left(1-2^{-(1-\alpha)}\right)$. The previous inequality chain and (16) imply $x^{-c} \geq \frac{1}{2}\psi_1$.

Case 2: $x^{-c} < (\log 2)/\lambda$. Then, by inequality (12),

$$1 - e^{-\lambda(1-\alpha)x^{-c}} \geq \frac{1}{2}(1-\alpha)\lambda x^{-c}. \tag{17}$$

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 (16) yields

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

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 7 *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 $x^c \in \left(x^c_{\max} - \varepsilon, x^c_{\max}\right)$.*

Proof. Suppose that $x^c \geq \psi_1/2$ and that $x^{-c} \notin \left(\psi_0\lambda^{-2}, \psi_1/2\right)$. Notice that from (13) and $x_{\max} = K\mu_c$ it follows that

$$\begin{aligned} x^c_{\max} - x^c &= K\mu_c - \left[K\mu_c\left(1-e^{-\lambda(1-\alpha)x^c}\right) + K\mu_{-c}\left(1-e^{-\lambda x^{-c}} - 1 + e^{-\lambda\alpha x^{-c}}\right)\right] \\ &= K\mu_c e^{-\lambda(1-\alpha)x^c} - K\mu_{-c}\left(e^{-\lambda x^{-c}} - e^{-\lambda\alpha x^{-c}}\right) \\ &= K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c}e^{-\lambda\alpha x^{-c}}\left(1-e^{-(1-\alpha)\lambda x^{-c}}\right). \end{aligned} \tag{18}$$

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 (18) 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 (18) 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 $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$. ■

Proof of Proposition 2. First, we show that for each $\varepsilon > 0$, there exists λ_0 such that for all $\lambda \geq \lambda_0$, for all equilibrium cutoffs x^c, either $x^c \leq \varepsilon$, or $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$. By Lemma 5, either $\max\{x^c, x^{-c}\} \leq \psi_0 \lambda^{-2}$ or $\max\{x^c, x^{-c}\} \geq \psi_1$. If $\max\{x^c, x^{-c}\} \leq \psi_0 \lambda^{-2}$ then $x^c \leq \varepsilon$ whenever $\lambda \geq \sqrt{\psi_0/\varepsilon}$. If $\max\{x^c, x^{-c}\} \geq \psi_1$, assume without loss of generality that $\max\{x^c, x^{-c}\} = x^c$. By Lemma 6, we have to consider only the following two cases: either $x^{-c} \leq \psi_0 \lambda^{-2}$, or $x^{-c} \geq (1/2)\psi_1$. If $x^{-c} \leq \psi_0 \lambda^{-2}$, for each ε there is a λ_0 such that if $\lambda \geq \lambda_0$ then $x^{-c} \leq \varepsilon$, and by Lemma 7, $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$. If $x^{-c} \geq (1/2)\psi_1$ then, since $x^c \geq \psi_1 > (1/2)\psi_1$, Lemma 7 implies that $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$ for $c \in \{b, w\}$.

Second, we show that equilibria of type (i)-(iv) exist. The colour-blind equilibrium is an example of an equilibrium of type (i). The existence of equilibria type (ii) and (iii) follows from Proposition 1. Finally, the proof of Lemma 7 implies that if $x^c \in (x^c_{\max} - \varepsilon, x^c_{\max})$ for each c, then $b^c(x^c, x^{-c}) \in (x^c_{\max} - \varepsilon, x^c_{\max})$ for each c. Since the functions b^c and b^{-c} are continuous and set $(x^c_{\max} - \varepsilon, x^c_{\max}) \times (x^{-c}_{\max} - \varepsilon, x^{-c}_{\max})$ is compact, the existence of equilibrium type (iv) follows from a standard fixed point argument. ■

References

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

Allport, G. W. (1979): *The Nature of Prejudice*. Basic Books.

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. Fryer (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.

Fryer, R. G. (2007): "Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century," *The Journal of Economic Perspectives*, 21(2), 71–90.

Gneezy, U., J. List, and M. K. Price (2012): "Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments," Working Paper 17855, National Bureau of Economic Research.

Haslam, S. (2004): *Psychology in Organizations*. Sage Publications, Thousand Oaks, CA.

Jones, S. R. G. (1984): *The Economics of Conformism*. Blackwell Pub.

Kandori, M. (1992): "Social Norms and Community Enforcement," *Review of Economic Studies*, 59, 63–80.

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 Inequality 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.

Minard, R. D. (1952): "Race Relationships in the Pocahontas Coal Field," *Journal of Social Issues*, 8(1), 29–44.

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

Peski, M., and B. Szentes (2012): "Spontaneous Discrimination," *SSRN eLibrary*.

Pettigrew, T. F. (1958): "Personality and Sociocultural Factors in Intergroup Attitudes: A Cross-National Comparison," *The Journal of Conflict Resolution*, 2(1), 29–42, ArticleType: research-article / Issue Title: Studies on Attitudes and Communication / Full publication date: Mar., 1958 / Copyright Âl' 1958 Sage Publications, Inc.

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

Pruthi, R. K. (2004): *Indian Caste System*. Discovery Publishing House.

Root, M. P. P. (2001): *Love's Revolution: Interracial Marriage*. Temple University Press.

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.

Sherif, M. (1961): *The Robbers Cave Experiment: Intergroup Conflict and Cooperation*. Wesleyan University Press.

Tajfel, H. (1970): "Experiments in Intergroup Discrimination," *Scientific American*, 223, 96–102.

Tajfel, H., M. Billig, R. P. Bundy, and C. Flament (1971): "Social Categorization and Intergroup Behaviour," *European Journal of Social Psychology*, 2, 149–178.

Tajfel, H., and J. Turner (1979): "An Integrative Theory of Intergroup Conflict," in *The Social Psychology of Intergroup Relations*, ed. by W. G. Austin, and S. Worchel. Brooks-Cole, Monterey, CA.