Spontaneous Discrimination[*]

Marcin Pęski[†]and Balázs Szentes[‡]

December 6, 2011

Abstract

This paper considers a dynamic economy in which agents are repeatedly matched with one another and decide whether 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 about the physical colour of agents in his partenrship 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, in some equilibria, agents condition their decisions on the physical and social colours of their potential partners. That is, they discriminate. The main result of the paper is that, in these aforementioned environments, every *stable* equilibrium must involve discrimination. In particular, the colour-blind equilibrium is unstable.

1 Introduction

Consider a town in the southern United States with a mix of non-white and white residents, some of whom are members of the Ku Klux Klan. These Klan members dislike their non-white neighbours, and are willing to punish those who even *associate* themselves with non-whites. The townspeople are otherwise tolerant, having no bias on the basis of skin colour. Each individual observes, perhaps imperfectly, social interactions within the community. Suppose now that a non-white community member is in search of employment. Klan members will obviously refuse to hire him. However, even an unbiased white employer, acting out of fear of punishment from the Klan, might refuse to hire a non-white applicant. In the end, this unfortunate job-seeker might face discrimination from the entire white community, and thus remain unemployed. Crucial in this story is the fact that individuals obtain some information about the interactions of others, for if hiring decisions weren't observed unbiased employers would be unafraid of hiring non-whites. But how many local

members does the Klan need in order to sustain discrimination? This paper shows that even if no one actually belongs to the Klan, the community might end up discriminating against non-whites.

This paper proposes a new theory of racial discrimination. The theory is based on the following assumption: each individual carries a label which conveys information about his past social interactions, and specifically the race of those he has chosen to associate himself with, directly or indirectly. Others are able to observe this label, and condition their decisions accordingly. One consequence, and indeed the main result of this paper, is that even individuals who are basically tolerant of other races might prefer to interact only with those of the same colour, and might also avoid those who even associate themselves with the opposite race. In other words, being labelled as someone who associates with one's own physical colour becomes valuable through the equilibrium play.

In an effort to develop the intuition behind this result, let us revisit the town described above and consider a white member of the community who holds no bias towards non-whites, but avoids interaction with them out of fear of punishment. He might also want nothing to do with anyone who has employed a non-white worker, because an indirect association with non-whites could be punished in the future. Taking this one step further, he might even avoid those who are only indirectly associated with non-whites despite never having employed them. As a result, employing non-whites and being associated with them is punished socially not only by the Ku Klux Klan, but by unbiased townspeople who are concerned about their reputations. This concern can be entirely self-enforcing and independent of the Klan; if a resident knows that others are reluctant to interact with anyone associated with non-whites, then he too is better off staying away from non-whites.

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 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.

We seek to characterize the *stable* equilibria of our model. An equilibrium is called stable if, after perturbing the equilibrium strategies slightly, myopic best-response dynamics imply convergence back to the equilibrium. The main result of this paper is that each stable equilibrium involves discrimination under certain conditions. In particular, the colour-blind equilibrium, in which agents ignore both physical and social colours, is unstable. Under our 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: one race

discriminates against the other, while members of the persecuted race use colour-blind strategies.

Let us emphasize that both the physical and social colours of the agents are payoff-irrelevant. Agents are motivated exclusively by monetary payoffs, and have no intrinsic preferences for one colour over another. Therefore, the equilibrium discrimination found in our model is not *taste-based*. In addition, agents of different colours are identical in terms of payoff-relevant characteristics, both ex-ante and ex-post. That is, an agent's physical colour reveals absolutely nothing about his potential as a business partner in terms of profitability. Therefore, the discrimination in our model is not statistical discrimination.

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)).

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. 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. Coate and Loury (1993) places Arrow's arguments in an equilibrium model but takes wages to be exogenous, as our model does. This assumption is relaxed in Moro and Norman (2004). Rosén (1997) offers another explanation for self-fulfilling statistical discrimination; workers observe their idiosyncratic productivity, privately, prior to applying for a job. If black workers choose to apply despite having a low productivity, firms rationally expect white applicants

3

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 observe the worker's productivities perfectly. However, the employers may decide not to search among black workers in anticipation of low skill-investment.

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.

Lang, Manove, and Dickens (2005) shows that even a slight presence of taste-based or statistical discrimination can have surprisingly large effects. Even if employers have only lexicographic preferences for white workers, or if white workers are only slightly more productive than black workers, the gap between white and black wages might be wide.

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.

In our model, a white agent discriminates against black workers because he fears that otherwise other white agents may refuse to hire him in the future. Punishment by peers for behaviour that differs from the accepted norm is a well-known phenomenon in sociology as well as in economics, see Austen-Smith and Fryer (2005) and the references therein.

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.[1] Because in our setup agents are risk-neutral, they have

---

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

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).

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. In this paper we also show that acting on payoff-irrelevant information is a possibility, but unlike Kandori (1992) and Mailath and Postlewaite (2006), we prove 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 arrive 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. 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 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

---

[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.

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 his past employment decisions, 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 history, his type, and the type of the worker into an employment decision. In what follows, we restrict our attention to *steady state equilibria*. That is, we characterize equilibria in which the agents' strategies depend neither on time nor on history.

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.

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

We fix a population strategy and a distribution of types at time zero. Neither the strategy nor the distribution need correspond to an equilibrium or be stationary. We derive the initial best-response cutoffs of each agent. 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 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 $x_{b,w}^w = \max\left\{0, \gamma\alpha\left(V_{w,w} - V_{w,b}\right)\right\}$. 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:*

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

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

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

Since agents of the same type have identical value functions, this lemma 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).

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}).

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

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

---

[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.

The rest of this section is devoted to 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 (3). This turns out to be useful when we later examine the stability properties of the equilibria. 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 the bias $V_{c,c} - V_{c,-c}$ for $c \in \{b, w\}$ given (X^b, X^w).[6] 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:

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

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

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

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 $\mu_{-c} (1 - EG(X^{-c}))$, where the expectation is taken according to the distribution of the cutoff X^{-c}. Finally, an employed worker's value changes to $V_{c,c}$, and he also receives 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,

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

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

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

That is,

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

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

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

where K denotes $M\delta\gamma/2r$.

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

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

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) if $b^c \left(x^c, x^{-c}\right) > 0$ then b^c is locally concave and strictly increasing in x^c,
  (ii) $b^c \left(0, x^{-c}\right) = 0$ for all x^{-c},
  (iii) for all $\overline{x}^{-c} > 0$, $b^c \left(x^c, 0\right) = \lim_{x^{-c} \to \infty} b^c \left(x^c, x^{-c}\right) \geq b^c \left(x^c, \overline{x}^{-c}\right)$.

Part (ii) implies that the function $b^c \left(x^c, 0\right)$ 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. The next lemma states that the existence of another intersection depends on the size of λ.

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 $x^c > 0$.

Figure 1 plots $b^b \left(., 0\right)$ and $b^b \left(., x^w\right)$ ($x^w > 0$) for the case when λ is large. The function $b^b \left(., 0\right)$ is identical to $\widetilde{b}^b \left(., 0\right)$ because $\widetilde{b}^b \left(., 0\right)$ is weakly positive (see (4) and (9)). For $x^w > 0$, $b^b \left(., x^w\right)$ is a downwards shift of $b^b \left(., 0\right)$, except it is zero whenever the shifted curve becomes negative. Since $\widetilde{b}^b$
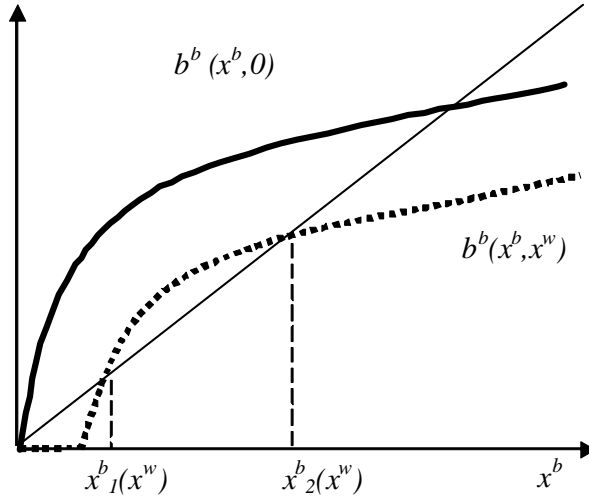
Figure 1: Best Responses

is concave in x^b, $b^b(., x^w)$ 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 $(x^b = 0)$ 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 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 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 $b^b(x^b, x^w)$ 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. 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. The slope of $b^b(., 0)$ is large if and only if λ is large (Lemma 4).

The positive intersection of $b^b(.,0)$ and the 45-degree line has an interesting interpretation. Suppose for a moment that white agents are non-strategic and their cutoff is zero, and consider our model as a game played by only black agents. If $b^b(x^b, 0) = x^b$ and $x^b > 0$, the best response of a black agent is x^b whenever every other black agent uses cutoff x^b. In other words, the cutoff x^c is an equilibrium in the game where only black agents act strategically. Since this cutoff is positive, black agents discriminate against others with white physical and social colour.

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 $\lambda > 1/(K(1-\alpha)\mu_c)$ for some $c \in \{b, w\}$. Second, we give a sharp characterization of the equilibrium strategies if λ is *very large*. In particular we prove that, in every equilibrium, the cutoff of an agent is either very small or very 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 $\lambda > 1/(K(1-\alpha)\mu_c)$, 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

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

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 (10).

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 (10) 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 is an $x^c > 0$ 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 (10). ∎

Note that in the proof of part (ii) we showed that if $b^c(x^c, 0) = x^c$, $x^c > 0$, then $(x^c, 0)$ is an equilibrium. Part (i) of Lemma 4 says that such an x^c does not only exist but is unique. Therefore, we can state the following

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

Part (ii) of Proposition 1 provides little information about the set of equilibria which involve discrimination. Although we do not characterize the set of equilibria when λ is large, we do establish some attributes of equilibrium strategies. In the next section, we will use these results to fully characterize all of the stable equilibria for the case of λ large.

Note that if (x_*^c, x_*^{-c}) is an equilibrium, then by (10), $x_*^c = b^c (x_*^c, x_*^{-c})$ 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 each x^{-c}, there are either two positive intersections of $b^c (., x^{-c})$ and the 45-degree line, or there are none. 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 show 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 $x^{-c} \notin (\underline{x}^{-c}, \overline{x}^{-c})$ and there is no intersection if $x^{-c} \in (\underline{x}^{-c}, \overline{x}^{-c})$. In this case, the curves x_1^c and x_2^c are only defined on $\mathbb{R}_+ \backslash [\underline{x}^{-c}, \overline{x}^{-c}]$. The next figure depicts x_1^c and x_2^c for both cases.
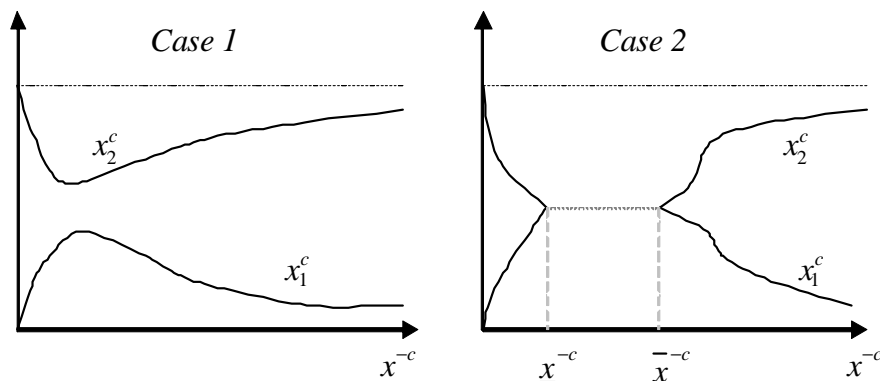

Figure 2: Positive Intersections

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. In the appendix, we 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 (iv) 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.[7] 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, $b^c(.,\underline{x}^{-c}) = b^c(.,\overline{x}^{-c})$. 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.

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

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

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

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

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

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 also define $x_i^c(0) = \lim_{x^{-c}\to 0} x_i^c(x^{-c})$. Note that $x_1^c(0) = 0$ and $x_2(0)$ corresponds to the positive intersection of $b^c(.,0)$ and the 45-degree line. In addition, 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 $\{x_i^b\}_{i=0}^2$ and $\{x_i^w\}_{i=0}^2$. Formally, (x_*^c,x_*^{-c}) is an equilibrium cutoff profile if and only if there exist $i,j \in \{0,1,2\}$ such that

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

Therefore, in order to find equilibria geometrically, we need to add the curves $\{x_i^{-c}\}_{i=0}^2$ 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.

---

[7]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.
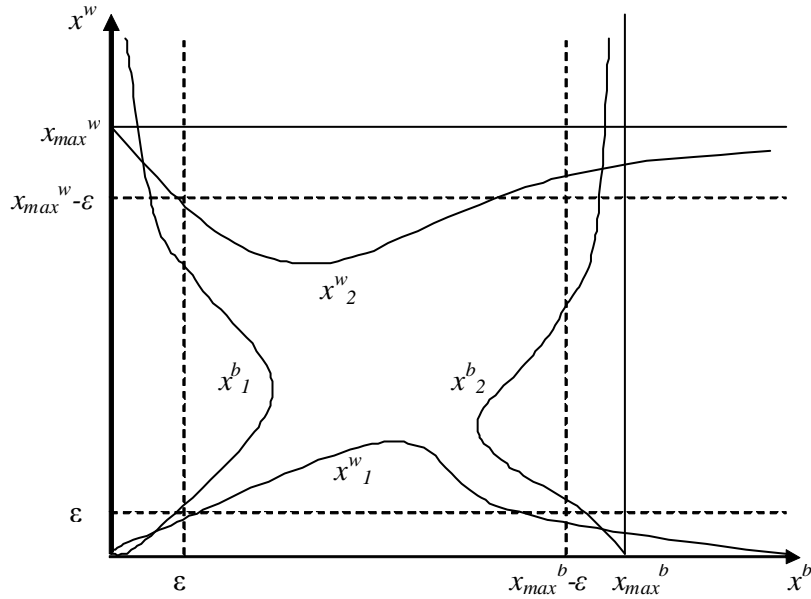
Figure 3: Equilibria

Note 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 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 ε (> 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 2 is illustrated in 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.

Note that if λ is large, then the realization of the shock is likely to be small. Hence, even if the cutoff of an employer is small, he might choose to not employ the worker with high probability. Can a weakly discriminating cutoff, $x^c_* \in (0, \varepsilon)$, generate high unemployment? In the proof of

15

Proposition 2, we show that it cannot. As λ goes to infinity, the probability of not employing induced by a weakly discriminating cutoff goes to zero.

The proposition allows any combination of these possibilities to be present in equilibrium. In Remark 1, we showed that the intersection of x_0^c and x_2^{-c} exists and is unique ($c \in \{b, w\}$). The unique intersection of x_0^c and x_0^{-c}, $(0,0)$, corresponds to the colour-blind equilibrium. Since $x_1^c(0) = 0$, the intersection of x_1^c and x_0^{-c} is $(0,0)$, that is, there is no equilibrium in which 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 with which we shall fully characterize those equilibria which are stable.

5 Stability

Next, we introduce a fairly standard notion of stability[8]. It is based on the requirement that a slight perturbation of agents' strategies around the equilibrium leads to the convergence of simple myopic best-response dynamics back to that equilibrium. We model the myopic best-response dynamics by assuming that each agent initially best-responds to some stationary population strategy. Each agent may be best-responding to a different population strategy which clearly might also differ from the strategy actually used by the population. Then, each agent stochastically receives an 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 were to never change.

Formally, suppose that black and white agents' initial cutoffs are denoted by a pair of random variables (X^b, X^w), and the strategy of each agent satisfies the statement of Lemma 1. This latter assumption is satisfied if each agent best responds to some population strategy. Agents receive opportunities to update their strategies according to a Poisson process with an arrival rate normalized to be one.[9] If an agent receives this opportunity at time t, he best-responds to the cutoff distribution at t as if it were constant over time. Let $x_t^c(X^c, X^{-c})$ 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 (x_*^b, x_*^w) 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 $\lim_{t \to \infty} x_t^c(X^c, X^{-c}) = x_*^c$ for $c \in \{b, w\}$.*

We next describe the equation that governs the best-response dynamics. Fix (X^b, X^w) and let (X_t^b, X_t^w) denote the population cutoffs at time t. We shall denote the myopic best response of an agent with colour c by x_t^c, suppressing its argument (X^b, X^w). By (8), the best-response of an

---

[8]See, for example, Chapter 3 of Fudenberg and Levine (1998).

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

agent with colour c at time t is

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

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 if he is able. 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,

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

where the first equality follows from (8). The second equality holds 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 third equality follows from (8) and (9). 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), (9) 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 illustrate the best-response dynamics derived from (12). Consider $x^b > x_2^b \left(x^w \right)$. 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 of the x_2^b curve, the best response of a black agent is below x^b. Equation (12) implies that in this region, the best response of a black agent decreases. This is represented by a horizontal arrow 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 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 now ready to state the main theorem of the paper.

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 $\left(x_*^w, 0 \right)$, $\left(0, x_*^b \right)$, and $\left(x_*'^w, x_*'^b \right)$ such that $x_*^c, x_*'^c \in \left(x_{\max}^c - \varepsilon, x_{\max}^c \right)$ for $c \in \{b, w\}$.*
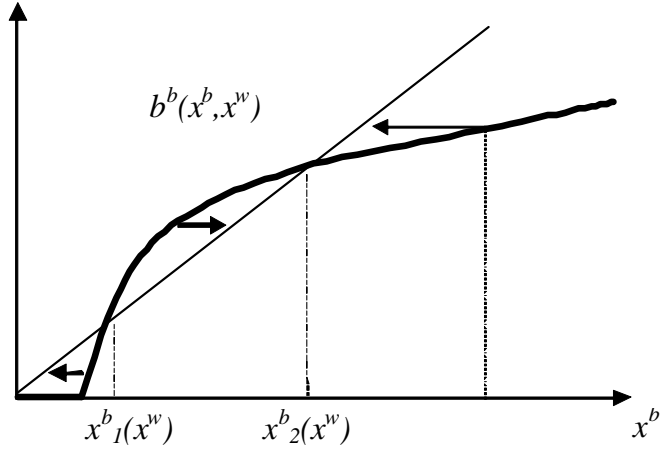
Figure 4: Best-response Dynamics

If λ is large enough, then our model gives rise to the following three stable equilibria. First, the white population discriminates strongly against the black population, while black agents do not discriminate at all. Second, the blacks discriminate strongly against the whites, while the whites do not discriminate at all. And finally, each race discriminates strongly against the other.

Equation (12) implies that the change in best responses at time t depends only on the time-t best responses themselves, $\left(x_t^b, x_t^w\right)$, but not directly on the distribution of strategies, $\left(X_t^b, X_t^w\right)$. In particular, the initial distribution of strategies affects the best-response dynamics only through the initial best-response profile $\left(x_0^b, x_0^w\right)$. Therefore, using (12), we can represent the best-response dynamics by constructing a Phase Diagram, plotted in Figure 5. For each cutoff vector $\left(x^b, x^w\right)$, the horizontal and vertical arrows on this figure indicate the directions of the best responses of black and white agents respectively.

To understand 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 indicated by an arrow pointing to the left. In general, if $\left(x^b, x^w\right)$ is to the right of the x_2^b 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 in the area that is to the right of the curve x_2^b. Similarly, Figure 3 shows that if $x_1^b\left(x^w\right) < x^b < x_2^b\left(x^w\right)$, then the best response of a black agent increases. This is why the horizontal arrows are pointing to the right between the curves x_1^b and x_2^b on Figure 5. Finally, if $x^b < x_1^b\left(x^w\right)$, then $b^b\left(x^b, x^w\right) < x^b$ which is indicated by horizontal arrows pointing to the left on the area that is to the left of x_1^b. The vertical arrows are constructed in a similar manner, representing the best-response dynamics of the white population.

Figure 5 can be useful in understanding 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 to the
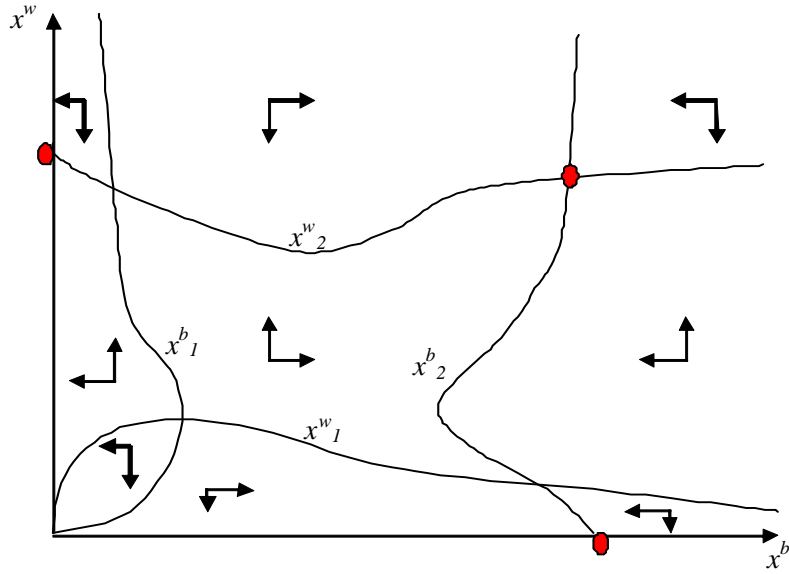
Figure 5: Stable Equilibria

right of 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.

Agents discriminate in equilibrium because social colour affects future employment. In particular, they are concerned about becoming workers and then being rejected by employers who discriminate. If λ is very small, then only a small fraction of an agent's payoff is due to the wage M. Therefore, agents care less about being workers and hence do not discriminate. How large does λ have to be for a version of Theorem 1 to hold? Recall that Proposition 1 states that the colour-blind equilibrium is the unique equilibrium if and only if the slope of the best response function is larger than one at the origin, that is, $\lambda \leq 1/\left(K\left(1-\alpha\right)\mu_c\right)$. From the proof of Theorem 1 it is clear that the colour-blind equilibrium is stable if and only if the slope of the best response curve is smaller than one at zero. Otherwise, each stable equilibrium involves discrimination. In fact, the equilibrium described in Remark 1 is stable. We formally state these results in the following

Theorem 2 *(i) If $\lambda \leq 1/\left(K\left(1-\alpha\right)\mu_c\right)$ for both $c \in \{b, w\}$, the equilibrium $(0,0)$ is stable.*

(ii) If $\lambda > 1/\left(K\left(1-\alpha\right)\mu_c\right)$ the profile $(0,0)$ is not stable and there exists a stable equilibrium.

6   Discussion

In this section, we begin by deriving some comparative static results. We then discuss some of the assumptions and extensions of our model. Finally, we 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 where λ is large, 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 in response to a shift in parameters. We emphasize that these comparative static results are valid only if the changes in the parameters are small enough that Theorem 1 still holds. The following table summarizes the comparative static results. This table shows what happens to the cutoffs if a certain parameter increases. It turns out that the direction of the change of 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 and is therefore omitted, but we will provide some intuition for these observations. An agent is concerned about his social colour because it holds influence over his future employment. Therefore, the more an agent's payoff depends on future wages, the more likely it is that employers condition their decisions on the types of the workers. For example, if the discount rate increases, agents care more about their current payoffs, relative to their future payoffs. Hence, they become more eager to employ workers of any type, and as a result, their cutoffs decrease. Similarly, if M increases, being an employed worker becomes more important, and the cutoffs increase. When λ increases, the expected payoff of an employer decreases, so having the option of being a worker becomes more important relative to being an employer. This is the reason that an increase in λ has the same effect as a decrease in M.

When μ_w increases, a larger fraction of agents' payoffs comes 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 remains unchanged with probability $1 - \gamma$ independently of his decision. So, the larger γ is, the more likely it is that the worker's type has an impact on the employer's future payoff. Hence, it becomes more important to discriminate.

6.2   Assumptions

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 the features of our model. Some of the assumptions we make are only necessary in order to provide a graphical representation of equilibria and stability. In what follows, we discuss some of our assumptions and extensions of the model.

Distribution of shocks.– We have assumed that the match-specific shock that determines the surplus of a partnership is exponentially distributed. 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 3 *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.

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 stability 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 stable equilibria different from those described in Theorem 1. 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. This can be motivated by the assumption 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 result regarding the instability of the colour-blind 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. The characterization of those processes which necessarily 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, a version of Theorem 3 will remain valid. That is, provided that λ is large, the colour-blind equilibrium will be unstable, and stable equilibria will exist. In other words, no matter how many attributes and labels are observable, adding 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 employed too infrequently. 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 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.

6.3 Empirical Implications

Our theory is different from taste-based and statistical theories of discrimination because agents have no intrinsic preferences for interacting with others of the same colour, and skin-colour provides

no signal about productivity. We next discuss empirical predictions of our model which are different from those of the other two theories.

It is not difficult to construct models of statistical or taste-based discrimination which generate the same comparative static results as the ones presented in Section 6.1. There are, however, testable implications of our theory which are unique. The key feature of our model is that the history of a worker affects the likelihood of his being hired in the present. A white employer, for example, uses a higher profitability-cutoff when deciding whether to hire a black worker than when he is faced with a white worker. 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, in our model 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 he interacted with more blacks in his 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, in our model, agents would never discriminate 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 are better able to observe the actions of others, than in larger communities, such as large cities, where individuals have less information about each other. This is in sharp contrast to the predictions of the other two theories.

One notable feature of our model is the presence of stable equilibria in which 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 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. Indeed, we showed that in some environments, every stable equilibrium must involve discrimination.

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.

8 Appendix

8.1 Proof of the Lemmas

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

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

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

(ii) By (9),

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

because $G\left(\alpha\overline{x}^{-c}\right) - G\left(\overline{x}^{-c}\right) \leq 0$. Hence, (4) and (9) imply $b^c(0, x^{-c}) = 0$.

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

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

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 $\widetilde{b}^c\left(x^c,0\right) \geq 0$ by (9), (4) implies $\widetilde{b}^c\left(x^c,0\right) = b^c\left(x^c,0\right)$. Therefore, by the proof of part (i) of Lemma 3

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

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 $\lambda > \lambda_0$ then $K\mu_c\left(1-\alpha\right)\lambda > 1$. This means that $\left.\partial b^c\left(x^c,0\right)/\partial x^c\right]_{x^c=0} > 1$ and therefore, $b^c\left(x^c,0\right) > x^c$ if x^c is close to 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 $\lambda \leq \lambda_0$ then $K\mu_c\left(1-\alpha\right)\lambda \leq 1$. 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. First, observe that by (9) and (4), $b^c\left(x^c,x^{-c}\right) - x^c$ if and only if $\widetilde{b}^c\left(x^c,x^{-c}\right) - x^c$. Therefore, we shall analyze the roots of the function $B^{x^{-c}}\left(x^c\right) \equiv \widetilde{b}^c\left(x^c,x^{-c}\right) - x^c$ for each x^{-c}. By (9)

$$B^{x^{-c}}\left(x^c\right) = K\mu_c G\left(\left(1-\alpha\right)x^c\right) - x^c + K\mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right).$$

Next, we establish some properties of $B^{x^{-c}}$ for $\lambda \geq 1/\left(K\left(1-\alpha\right)\mu_c\right)$.

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

(2) $\left.dB^{x^{-c}}/dx^c\right|_{x_c=0} > 0$. 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).[10]

Note that the part of $B^{x^{-c}}\left(x^c\right)$ which depends on x^{-c}, $K\mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)$, is additively separable. Hence, the curve $B^{x^{-c}}\left(x^c\right)$ is a vertical shift of $B^0\left(x^c\right)$. The number of roots of $B^{x^{-c}}\left(x^c\right)$ depends on the size of this shift. Let $H\left(x^{-c}\right)$ denote this shift, that is, $H\left(x^{-c}\right) = K\mu_{-c}\left(G\left(\alpha x^{-c}\right) - G\left(x^{-c}\right)\right)$. The following properties of H are straightforward consequences of $G\left(s\right) = 1 - e^{\lambda s}$:

(6) $H\left(x^{-c}\right) < 0$ if $x^{-c} > 0$.

(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 lemma.

(i) Suppose that $\max_{x^{-c}} B^{\widehat{x}^{-c}}\left(x^c\right) > 0$. This, together with (8), implies that $\max_{x^{-c}} B^{x^{-c}}\left(x^c\right) > 0$ for all $x^{-c} > 0$. On the other hand, $\inf_{x^{-c}} B^{x^{-c}}\left(x^c\right) < 0$ by (3). Hence, the Intermediate Value Theorem implies that $B^{x^{-c}}$ has at least one root. Therefore, by (5), $B^{x^{-c}}$ has exactly two roots; $x_1^c\left(x^{-c}\right)$ and $x_2^c\left(x^{-c}\right)$.

(ii) Suppose that $\max B^{\widehat{x}^{-c}}\left(x^c\right) < 0$. 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

---

[10] There is a non-generic case when $B^{x^{-c}}$ is tangent to the constant zero line.

by $(\underline{x}^{-c}, \overline{x}^{-c})$. From (2) and $B^0(0) = 0$ it follows that $\underline{x}^{-c} > 0$. From (7) it follows $\overline{x}^{-c} < \infty$. If $x^{-c} \in (\underline{x}^{-c}, \overline{x}^{-c})$, $B^{x^{-c}} < 0$ and it has no roots. If $x^{-c} \in \mathbb{R}_+ \backslash [\underline{x}^{-c}, \overline{x}^{-c}]$ then the same argument as in part (i) shows that $B^{x^{-c}}$ has two roots, $x_1^c(x^{-c})$ and $x_2^c(x^{-c})$. The values $\underline{x}^{-c}$ and $\overline{x}^{-c}$ correspond to the non-generic case where $B^{x^{-c}}$ is tangent to constant zero line. From (1) and (9) it follows that

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

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

8.2 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

$$\psi_0 = \frac{1}{\frac{1}{4} K \alpha (1 - \alpha) \mu_{\min}}, \tag{13}$$

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

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

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

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

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),

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

26

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 $\max\{x^b, x^w\} \geq (\log 2)/\lambda$, then from the previous equality it follows that

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

Since $\max\{x^b, x^w\} \geq \frac{1}{2}(x^b + x^w)$, the previous inequality chain implies $\max\{x^b, x^w\} \geq \psi_1$. If $\max\{x^b, x^w\} \leq (\log 2)/\lambda$, then, by Lemma 6,

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

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

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

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

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

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

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

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

If $\lambda > 1/(2K\mu_w(1-\alpha))$ 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 $x^c \geq \psi_1$. Suppose that $x^{-c} > 0$. Then

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

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

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

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

$$\geq K\mu_{-c}\left(1 - 2^{-(1-\alpha)}\right) - \frac{1}{2}\psi_1 \geq \frac{1}{2}\psi_1,$$

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: $x^{-c} < (\log 2)/\lambda$. 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

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

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 $x^c \in (x_{\max}^c - \varepsilon, x_{\max}^c)$.*

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

$$x_{\max}^c - x^c = K\mu_c - \left[K\mu_c\left(1 - e^{-\lambda(1-\alpha)x^c}\right) + K\mu_{-c}\left(1 - e^{-\lambda x^{-c}} - 1 + e^{-\lambda\alpha x^{-c}}\right)\right] \tag{19}$$

$$= K\mu_c e^{-\lambda(1-\alpha)x^c} - K\mu_{-c}\left(e^{-\lambda x^{-c}} - e^{-\lambda\alpha x^{-c}}\right)$$

$$= K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c}e^{-\lambda\alpha x^{-c}}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right).$$

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

$$K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c}e^{-\lambda\alpha x^{-c}}\left(1 - e^{-(1-\alpha)\lambda x^{-c}}\right) \leq K\mu_c e^{-\lambda(1-\alpha)x^c} + K\mu_{-c}e^{-\lambda\alpha x^{-c}}$$

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

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

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

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, $x^c \in (x_{\max}^c - \varepsilon, x_{\max}^c)$.

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

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

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 $x^c \in (x_{\max}^c - \varepsilon, x_{\max}^c)$. ∎

We are ready to prove Proposition 2. 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 2.

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, $x^c \in (x_{\max}^c - \varepsilon, x_{\max}^c)$. 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 $x^c \in (x_{\max}^c - \varepsilon, x_{\max}^c)$ for $c \in \{b, w\}$. Therefore, (iii) of the statement of Proposition 2 holds for $c \in \{b, w\}$.

8.3   Proof of Theorem 1

By Proposition 2, 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	$\in (x_{\max}^{-c} - \varepsilon, x_{\max}^{-c})$
Case 4	$\in (0, \varepsilon)$	$\in (0, \varepsilon)$
Case 5	$\in (0, \varepsilon)$	$\in (x_{\max}^{-c} - \varepsilon, x_{\max}^{-c})$
Case 6	$\in (x_{\max}^{-c} - \varepsilon, x_{\max}^{-c})$	$\in (x_{\max}^{-c} - \varepsilon, x_{\max}^{-c})$

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 $\delta \in (0, \log 2/[\lambda(1-\alpha)])$. Let the best response of an agent with colour c at time t denoted by x_t^c if the initial distribution of cutoffs is (X^c, X^{-c}). Equations (8) and (12) imply that $x_t^{-c} = 0$ for all t. However, we show that x_t^c does not converge to 0 for sufficiently large λ. Since $x_0^c > 0$, it is enough to prove that $dx_t^c/dt > 0$ whenever x_t^c is small but positive. Suppose that $x_t^c \in (0, \log 2/[\lambda(1-\alpha)])$. Then

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

where the first equality is just (12) with $x_t^{-c} = 0$ and the inequality follows from $x_t^c \in (0, \log 2/[\lambda(1-\alpha)])$ 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 2 and Lemma 7 imply that $x^c \in \left(0, \psi_0 \lambda^{-2}\right)$ if λ is large enough. In Case 5, Proposition 2 and Lemma 8 imply that $x^c \in \left(0, \psi_0 \lambda^{-2}\right)$ if λ is large enough. Also notice that

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

and for sufficiently large λ,

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

$$\geq \frac{1}{2}K\mu_c\lambda\left(1 - \alpha\right),$$

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,

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

Case 3.

Remark 1 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 $\left|x^{-c} - x_2^{-c}\left(x^c\right)\right| < \delta_2$. In addition, we established in Section 5 that if δ_1 and δ_2 is small enough, $x^c < \delta_1$ and $\left|x^{-c} - x_2^{-c}(0)\right| < \delta_2$ then

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

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

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

for all t.[11] 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.

---

[11] 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 $x_t^{-c} = x_2^{-c}(0) - \delta_2$.

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 $\lambda > \lambda_0$,

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

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), $\left(x_*^c, x_*^{-c}\right)$ is an equilibrium cutoff profile.

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 $\left(x^c, x^{-c}\right)$ 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 $\left(x^c, x^{-c}\right)$ 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. 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.

FUDENBERG, D., AND D. K. LEVINE (1998): *The Theory of Learning in Games*. The MIT Press.

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 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.