Discrimination and Workers' Expectations

Antonio Filippin
University of Milan and IZA
antonio.filippin@unimi.it

Preliminary version
Comments welcome
Do not quote without permission

March 30, 2005

Abstract

The paper explores the role of workers' expectations as an original explanation for the puzzling long run persistence of observed discrimination against some minorities in the labor market. A game of incomplete information is presented, showing that *ex ante* identical groups of workers may be characterized by unequal outcomes in equilibrium due to their different beliefs, even though discriminatory tastes have disappeared. Wrong beliefs of being discriminated against are self-confirming in this circumstance, being the ultimate cause of a lower percentage of promotions which supports these wrong beliefs. Unequal outcomes are rationalized by employers with a statistical discrimination argument, which however does not have behavioral implications and therefore does not affect the distribution of promotions. Unequal outcomes may also be rationalized by majority workers via self-serving beliefs about the distribution of ability across populations. Unequal outcomes driven by workers' expectations turn out to be robust to trial work periods and to affirmative actions like quotas.

Keywords: Discrimination, Workers' Expectations, Self-Confirming Beliefs.

JEL classification codes: J71, J15, J24, D84, C79.

Address correspondence to Antonio Filippin, University of Milan, Department of Economics, via Conservatorio 7, 20122 Milano, Italy. I would like to thank and Pierpaolo Battigalli, Thomas Bauer, Giuseppe Bertola, Roger Farmer, Andrea Ichino, Ramon Marimon, Massimo Motta, Giulana Palumbo, as well as seminar participants at EEA 2002, IZA Summer School 2002, RES 2003, AIEL 2003, 1st Brucchi Luchino Workshop, EUI, University of Milan, University of Turin, and University of Modena for their valuable suggestions. Needless to say, all the remaining errors are mine.

1 Introduction

Despite several contributions to the literature, there is still no widely shared explanation for the long-run persistence of discrimination in the labor markets. Moreover, the neoclassical theory of discrimination is mostly a demand-side theory. There are very few contributions where workers' heterogeneity matters, and, to the best of my knowledge, only a recent paper by Breen and Garcia-Penalosa (2002) studies the possibility that unequal outcomes may persist for reasons that can be ascribed to workers' expectations.

There is instead empirical literature that points toward an important role played by workers' expectations of being discriminated against. For instance, Filippin (2003b) and Filippin (2004) present experimental evidence that subjects belonging to populations randomly generated in the lab (red Vs. blue) are likely to show a lower propensity to bid in an all-pay auction after having been discriminated against for a while when the prize of the auction is awarded. Even more interestingly, Hoff and Pandey (2004) conducted a field experiment recruiting children belonging to different Indian castes with the task of solving mazes. There were no caste differences in performance when caste was not publicly revealed, but making caste salient created a large and robust caste gap as long as there was scope for discretion and judgment in rewarding performance.

The goal of this paper is to analyze from the theoretical point of view the role of workers' expectations, so far neglected in the literature, in explaining the observed unequal outcomes that characterize some minorities in the labor market. The idea is that minority groups who expect being discriminated against exert a lower effort on average, because of a lower expected return. This induces a lower percentage of promotions within minority workers even though employers do not discriminate against them. In turn, this outcome is consistent with minority workers' beliefs that there are employers characterized by discriminatory tastes.

The model is formalized as a game of incomplete information in which populations of workers and employers are engaged. In every constituent game, i.e. in every repetition of the game played by actors randomly drawn from their populations, three players participate: one employer and two workers, one of whom belongs to a minority group. The employer promotes one (and only one) of the two workers after having observed their effort. Crucially for the results of this chapter, promotions depend via effort on workers' expectations about the unknown employer's type, which captures his possible disutility of promoting a minority worker.

The importance of workers' expectations can be appreciated comparing the equilibrium outcome in terms of promotions arising when minority workers overestimate the percentage of discriminatory employers as compared to a situation in which such beliefs are correct *ceteris paribus*. Even in a labor market where employers do not discriminate against minority workers, and where the distribution of ability is the same across groups of workers, unequal outcomes may still arise due only to workers' expectations. It is worth stressing that such assumptions are made in order to test workers' expectations as a "stand-alone" source of unequal outcomes from a theoretical point of view, not because other sources are regarded as negligible. What happens is that wrong beliefs of being discriminated against are self-confirming in equilibrium.

Unequal outcomes are rationalized by the employers using a statistical discrimination argument, which however does not have behavioral implications and therefore does not affect the distribution of promotions. On the contrary, unbiased employers implement a fair tournament. Unequal outcomes may also be rationalized by majority workers by means of self-serving beliefs concerning the distribution of ability across populations, i.e. overestimating their individual as well as population ability in a way that enhances their self-esteem.

The result that that unequal outcomes can be ascribed to workers' different expectations, is robust both to trial work periods, which are instead an effective policy device to break down statistical discrimination outcomes, and to affirmative actions like quotas. The conclusion is that workers' expectations can contribute to explain why historically oppressed social groups are not likely to forge ahead once the original cause of unequal outcomes has been removed.

The model in this paper also presents some nice features from the methodological point of view. To my knowledge, it is one of the few applications that relaxes the common prior assumption in a fruitful way to go on an applied problem. The violation of the Common Prior assumption has the strong implication that beliefs in a Perfect Bayesian Equilibrium may be contradicted by the evidence, in such a way that this commonly used equilibrium concept could not be associated with a fixed point of a learning process. For this reason, the equilibria presented in this paper are also self-confirming, i.e. not contradicted by the evidence, so that they can safely be interpreted as a fixed point of a learning process. Moreover, the common prior assumption is relaxed even though players are given access to the same information. In other words, different beliefs are not justified by different learning processes. On the contrary, workers interpret the same aggregate observables in a different, and self-serving, way.[1]

The structure of the paper is as follows. After some definitions (Section 2.1) and the connections of the model to the related literature (Section 2.2) are sketched, the constituent game of the model, i.e. the game after the players have already been matched, is presented (Section 3.1). The population game, the matching process and the information structure, necessary to characterize beliefs, are described in Section 3.2. Section 4 concentrates on the analysis of the equilibria of the model and its policy implications. Section 5 concludes.

[1] This is in line with the evidence emerging from a sample of second year students of Bocconi University, as reported by Filippin and Ichino (2005), that while more males than females think that actual differences between men and women matter, a larger fraction of females points toward employers discriminatory tastes as one of the causes for the expected gender gap, the size of which is expected to be similar by men and women.

2 Definitions and Related Literature

Before starting the presentation of the model, it is useful to clarify the meaning attached to some concepts throughout the paper and to analyze the theoretical literature concerning discrimination.

2.1 Definitions

In the literature many different and occasionally contradicting definitions have been used referring to discrimination in the labor market. Discrimination has been defined either as different achievements (wages, promotions) for equally *productive* workers, or as different achievements for workers that have the same characteristics *ex ante*, i.e. for workers with the same *ability* and *taste for work*. Not infrequently, the two concepts have been used interchangeably, but this seems inappropriate, because *ex ante* equal workers can be characterized by different productivity in equilibrium.

A good compromise, partially following Blau, Ferber and Winkler (2002), is to use two different definitions. On the one hand, following the "equal pay for equal work" principle, **direct discrimination** can be defined as a different treatment in terms of wages, promotions, or job allocations for equally productive workers.[2] On the other hand, a more comprehensive definition seems to be necessary, too. The reason is that it would be hard to consider as discriminatory an employer who pays or promotes minority workers less (on average) if they are (on average) proportionally less productive. Nevertheless, it would be misleading to disregard the fact that many factors, and direct discrimination can be one of the most important, may affect workers' behavior. If minority workers are less productive, for example, because they have changed their behavior reacting to a worse job assignment, the different achievements should not be viewed as equal treatment, even if there is no evidence of direct discrimination. Such a situation is captured by the more comprehensive concept of **cumulative discrimination**, defined as different achievements for workers that have the same characteristics *ex ante*.

Another distinction that deserves to be mentioned is that between *group* and *individual* discrimination. The former happens when different achievements are observed on average either between groups of workers that are on average equally productive (direct group discrimination) or between groups of workers which are *ex ante* equal (cumulative group discrimination). The latter happens when an individual is judged on the basis of group membership rather than upon his or her own characteristics only. Individual discrimination is a characteristic of all the models of incomplete information and concerns both the majority and the minority group. Moreover, it does not imply group discrimination. Henceforth, whenever not specified, discrimination always refers to cumulative group discrimination.

[2] Often the definition of direct discrimination refers to "equally qualified" workers.

2.2 Related Literature

The model presented in Section 3 can be fruitfully compared with most of the contributions to the discrimination literature.[3] One thing that must be taken into account is that although these contributions often focus on wages rather than promotions, the main stylized facts can be replicated focusing on promotions as well.

Six groups of models are presented: discriminatory tastes, statistical discrimination, human capital theory, feedback effects, workers' expectations and asymmetric tournaments.

Discriminatory Tastes. The starting point of the economic analysis of discrimination in labor markets can be found in the article "The Economics of Discrimination" by Becker (1957). In Becker's model, the existence of *direct discrimination* between workers of different groups, which are perfect substitutes in the production function, is based on the discriminatory preferences of employers, coworkers or customers. Hence, discrimination is caused by fundamentals (discriminatory tastes), while beliefs do not play any role because there is no uncertainty. Within this framework, members of the discriminated group must receive a lower wage in order to be accepted as employees, coworkers or sales. Among the advantages of Becker's approach, there is the possibility of explaining the rise of any type of direct discrimination (based on sex, race, religion, etc.). On the other hand, the major problem lies in its long run implications: if markets are competitive and there is heterogeneity of discriminatory tastes, only the less discriminatory employers (or the non-discriminatory ones if present) should survive. The reason is that discrimination is costly for the employer, so that when competition drives profits toward zero discriminatory employers would suffer a negative utility. Alternatively, we should observe complete segregation. However, both predictions are contradicted by empirical evidence.

Statistical Discrimination. In the statistical discrimination models, group membership is assumed to convey information regarding individual characteristics, about which incomplete information is assumed. Several models have been developed within this strand of literature, using different devices in order to explain the long-run persistence of observed discrimination. Common to these models is the fact that, unlike Becker's one, fundamentals are not relevant.

The seminal contribution in the statistical discrimination literature has been proposed by Arrow (1973).[4] Employer's beliefs about the existence of different

[3] In this section, theories have been selected and outlined in such a way as to facilitate contrast and comparison with the model of workers' expectations. Therefore, the choice of the contributions to be summarized is far from exhaustive, focusing only on the theoretical aspects of some competitive neoclassical models and institutional theories. Also the relative weights assigned to various aspects of such theories reflect primarily the necessity of the subsequent presentation, rather than some sort of consensus about what has been considered more important in the literature thus far. Another reason for these choices is that many detailed surveys are already available (see Blau, Ferber and Winkler (2002) and Cain (1986) among others).

[4] Other examples of statistical discrimination can be found in Phelps (1972), who concentrates on the effect of an imperfect predictor of the true productivity of a worker, and

characteristics between (*ex ante* identical) groups turn out to be correct in equilibrium.[5] Why are these expectations confirmed in equilibrium? In other words, why are these wrong beliefs self-confirming? The mechanism is the following: a worker's a priori unobservable variable (e.g. effort) is endogenously affected by employer's beliefs (e.g. via lower wages, or via worse job assignments), leading to a suboptimal investment in her skills (or a suboptimal supply of effort) and therefore determining an outcome that confirms the beliefs of the employer. The conclusion is that in equilibrium there is *cumulative but not direct discrimination*, because workers are *ex ante* equal but show a different productivity in equilibrium.

Statistical discrimination models have been criticized by Cain (1986), on the ground that "these models face the criticism that the employer's uncertainty about the productivity of workers may be inexpensively reduced by observing the workers' on-the-job performance." Workers' performance can be observed for example by means of trial work periods. Cain's argument are straightforwardly encompassed into the model presented in this paper, where updated beliefs are used to decide on promotions and where the whole first period can be thought as a trial work period. Nonetheless, the statistical discrimination model plus trial work period leaves some open questions: what determines workers' behavior in the trial work period? Is it convenient for them to increase effort to be assigned to the good job? The answers to these questions cannot be found within the statistical discrimination literature, because it is necessary to analyze also the supply side of the labor market. In section 4, where the role of workers' expectations is analyzed, it emerges that trial work periods are not an effective policy device to break down unequal outcomes, as long as minority workers believe they are discriminated against.

The Human Capital Theory. Another strand of the literature, started by Mincer and Polacheck (1974), is the so-called human capital theory which analyzes the effects of voluntary choices of investment in human capital from a gender perspective. According to this theory, women decide to invest less than men because they expect a lower lifetime return on human capital due to a shorter and more discontinuous presence in the labor force. As a consequence, they receive less on-the-job training and/or are assigned to less rewarding jobs. Such behavior can be ascribed to the traditional division of work within the family (Becker, 1985). In this way, wage differentials, worse career path, and/or sex segregation are explained by voluntary choices. If this is the case, the different achievements *could not be classified as discrimination*, given that workers neither equally productive in equilibrium nor *ex ante* equal.

Some economists have heavily criticized this approach (see the next paragraph), because this seemingly "voluntary" decision could actually be induced

Spence (1973), in his pioneering work about signaling. A skeptical reading of the statistical discrimination approach can be found in Aigner and Cain (1977) and Cain (1986). Some of the arguments raised by Cain are also relevant in the model of workers' expectation presented in this paper and have been addressed in section 2.3.4.

[5] Moro and Norman (2002) analyze statistical discrimination using a general equilibrium approach.

by discrimination, entering the definition of cumulative discrimination.

Feedback Effects. The boundaries of this approach are particularly uncertain,[6] and usually surveys concerning labor market discrimination use these models as a counterpart for other theories, without analyzing them separately. The reason is that the contributions that can be grouped into this category are quite heterogeneous: the main idea they have in common is that the behavior of the workers can in turn be determined by discrimination. However, the mechanisms through which the behavior is affected vary considerably. In many cases there is also a lack of formalization and these effects are little more than qualitative statements.

Blau and Jusenius (1976), reverse the causality link with respect to Mincer and Polacheck (1974): women, because of experiences of sex discrimination, e.g. lower wages, respond with career interruptions and specialization in household production, i.e. investing less in human capital.

The explicit analysis of workers' expectation in this paper is a way to formalize such feedback effects.

Workers' expectations. As already mentioned, the neoclassical theory of discrimination is mostly a demand-side theory. But why should workers' preferences not be allowed to play a role as important as that of either employers' preference in the discriminatory tastes approach or employer's beliefs in the statistical discrimination models?

To the best of my knowledge, the only paper in the literature on discrimination that focuses on the supply side of the labor market is that of Breen and Garcia-Penalosa (2002), who explain the observed persistence of gender segregation using a Bayesian learning approach. Workers, due to imperfect information, do not know and try to learn how much the probability of success in various occupations is affected by effort or by predetermined individual characteristics (such as gender). The "prior" of a man (woman) is the belief received by his father (her mother), while the posterior is the belief updated according to his (her) experience and transmitted to his son (her daughter). Different preferences between men and women at some point in the past caused different learning paths and different beliefs. This is a sufficient condition to observe lasting unequal outcomes in equilibrium for the two groups, even once preferences become equal, meaning that past circumstances continue to exert an influence and that expectations can be self-fulfilling.

Similarities of this paper with the work of Breen and Garcia-Penalosa are evident: both consider the effect of heterogeneity within the supply side of

[6] A large number of the so called "institutional" contributions may also fall into this category. Cain (1986) includes also the above-mentioned model by Arrow (1973) within this group. The seminal "institutional" contribution has been made by Myrdal (1944), who theorizes the "principle of cumulation," a mechanism of dynamic causation between several variables. These variables move together influencing each other once the system is hit by an external shock. Among the secondary causes of discrimination, the behavior of workers is also taken explicitly into account: "Negro worker often feels that his fate depends less on his individual efforts than on what white people believe about Negroes in general" (Myrdal, 1944). Other contributions follow along the line of the vicious circle described by Myrdal, like Ferber and Lowry (1976).

the labor market and both explain the persistence of unequal outcomes via self-confirming workers' expectations. What differs in the model of Breen and Garcia-Penalosa is a different information structure. Agents learn from their parents only, but not from observable aggregate outcomes. Moreover, only agents choosing a "high" profile of education and effort are able to learn from their experience and transmit updated beliefs to their children, while for the "low" profile the learning process stops. The key mechanism behind the results of these authors, is that the information structure of the model prevents agents from learning that differences in fundamentals have disappeared. In other words, beliefs are still a function of differences in workers' fundamentals. Section 4 will show that workers' expectations can explain observed unequal outcomes even when such an assumption is relaxed because all workers have the possibility of learning from the same aggregate observables.

Asymmetric Tournaments. A tournament is symmetric when outcomes are invariant to the permutation of the contestants. On the other hand, asymmetric contests are defined "uneven" when agents are different, and "unfair" when contestants are identical but the rules favor one of them. The literature on tournaments, started by Lazear and Rosen (1981), is not directly related to discrimination. Nevertheless, asymmetric tournaments, as described by O'Keeffe, Viscusi, and Zeckhauser (1984), provide a useful framework for the analysis of the effects of discrimination on promotions and are therefore a useful benchmark for the game presented in this chapter. Not surprisingly, the two models provide similar predictions in some cases, e.g. that discriminatory tastes, as an example of unfair rules, affect the incentives of both minority and majority workers in the same way.

Within the literature on uneven tournaments, it is incidentally mentioned that unequal outcomes may arise between groups of workers that are *ex ante* equal.[7] However, the underlying mechanism has not been formalized and, more specifically, no role is explicitly played by expectations. The result, presented in Section 4 that effort differs across otherwise identical workers because of their different beliefs, may also be interpreted as a formal justification for the existence of uneven tournaments between *ex ante* equal workers.

3 The Model

The model is formalized as a two-stage game of incomplete information where populations of workers and employers are engaged. The two populations of workers differ because of an observable characteristic (race, gender, etc.) which does not affect their output (productivity) π. The observable characteristic distinguishes the so called majority worker, identified by subscript A, from the so called minority worker, identified by subscript B. Employers are denoted by subscript F. Workers compete in order to be promoted. Promotions depend on both employers' and workers' type as well as on their beliefs about the oppo-

[7] See Schotter and Weigelt (1992).

nents' type-strategy profile. Crucially for the results of this chapter, promotions depend via effort on workers' expectations about the unknown employer's type, which captures his possible disutility of promoting a minority worker.

The following section focuses on the constituent game, i.e. on what happens after the players have been drawn from their populations and matched. The population game, the matching process and the information structure, necessary to characterize beliefs, are described in Section 3.2.

3.1 The constituent game

In every constituent game two workers, one of which is a "minority" worker, and one employer are drawn from their populations and play a two-stage game. In the first period both workers choose a level of effort. The employer observes workers' productivity in the first period and promotes one (and only one) of the two workers. After having observed employer's decision, workers choose again a level of effort in the second period.

Assumption 1.

a) *Labor market is competitive*, therefore productivity is entirely paid to workers ($w = \pi$). This assumption makes the game equivalent to the reduced form of a more general game where workers' output is observed and verifiable and employers compete on enforceable piece-rate contracts. Workers are free to move, but in equilibrium $w = \pi$ and nobody moves.

b) *Productivity and effort:* productivity coincides with effort in the first period and for the non-promoted worker in the second, while it is given by the interaction between effort and a function of ability in the second:[8]

$$\pi_A^1 = e_A^1; \quad \pi_A^2(\alpha_A) = \sqrt{\theta_A} e_A^2; \quad \pi_A^2(\alpha_B) = e_A^2$$
$$\pi_B^1 = e_B^1; \quad \pi_B^2(\alpha_B) = \sqrt{\theta_B} e_B^2; \quad \pi_B^2(\alpha_A) = e_B^2$$

where α_A means that worker A is promoted, α_B that worker B is promoted, and the superscripts 1 and 2 the periods.

The constituent game is characterized by observable actions, because the decision about promotion is directly observed and every level of (observed) output is one to one related with effort.

3.1.1 Incomplete Information

In this game every player knows his/her own type only.

$\theta_A \in [1, 2]$ and $\theta_B \in [1, 2]$ summarize the ability of majority and minority workers, respectively.

$\theta_F \in \Theta_F$ represents employer's tastes for discrimination. There are only two types of employer $\Theta_F = \{0, d\}$. If $\theta_F = 0$ (henceforth: θ_0) the employer is indifferent about the observable characteristic which distinguishes the workers. On the other hand, if $\theta_F = d$ (henceforth: θ_d) the employer suffers a disutility

[8]The square root of ability is used just to simplify as much as possible the expressions for the optimal level of effort

when the minority worker is promoted. The disutility d is assumed to be sufficiently high that promoting worker A is always the optimal choice regardless of workers' productivity. Without loss of generality, it is assumed that $Pr(\theta_0) = 1$, i.e. that there are only unbiased employers.

Summarizing, a minority worker knows her own ability θ_B, while the type θ_B of the majority worker and the tastes for discrimination θ_F of the employer are unknown.[9]

3.1.2 Payoffs

The structure of the utility function is the same for majority (A) and minority (B) workers and it is parametrized according to their type θ:[10]

$$U^{\theta_P} = w_P^1 - \frac{(e_P^1)^2}{K_1} + w_P^2 - \frac{(e_P^2)^2}{K_2},$$

where:

w_P^t is the wage in period t for the worker belonging to population $P = (A, B)$;

e_P^t is effort in period t for the worker belonging to population $P = (A, B)$;

K_1, K_2 are two constant that weight the disutility of effort in the two periods. Although nothing prevents the two constants from being equal in the two periods, and even from being equal to one, I assume that $K_1 = 2$ and $K_2 = 4$ in order to obtain simpler equations for the optimal choice of effort. Substituting $K_1 = 2$ and $K_2 = 4$ and exploiting assumption 1, we get:

$$U^{\theta_P} = e_P^1 - \frac{(e_P^1)^2}{2} + I(\alpha_P)\left[\sqrt{\theta_P}e_P^2 - \frac{(e_P^2)^2}{4}\right] + I(\alpha_{-P})\left[(e_P^2) - \frac{(e_P^2)^2}{4}\right], \quad (1)$$

where:

U^{θ_P} represents the utility of type θ of population P;

$I(\cdot)$ is the indicator function that assigns the value 1 when the argument is true, for instance when the worker P is promoted $I(\alpha_P) = 1, I(\alpha_{-P}) = 0$.

Using equation (1) it is straightforward to compute that in the second period effort will be higher if the worker is promoted, and in particular:

$$e_P^{2*}|\alpha_P = 2\sqrt{\theta_P} \geq e_P^{2*}|\alpha_P = 2.$$

Applying backward induction and exploiting Assumption 1, it is possible to write the expected utility as a function of the effort exerted in the first period

$$EU^{\theta_P}(e_P^1) = e_P^1 - \frac{(e_P^1)^2}{2} + 1 + \mu_P(\alpha_P(e_P^1))(\theta_P - 1), \quad (2)$$

where:

[9] Of course, the same holds *mutatis mutandis* for players B and F.

[10] The specification of the utility function adopted in the paper is the same that have been proposed by the asymmetric tournament literature (see O'Keeffe, Viscusi, and Zeckhauser (1984)). The only difference is that here the role of the prize is played by the higher wage attached to the job of the promoted worker.

$\theta_P - 1$ is the gain of utility when promoted.

$\mu_P(\alpha_P(e_P^1))$ represents the subjective probability of being promoted, as a function of the effort exerted in the first period.

As far as the employer is concerned, the utility function contains both profits and a parameter summarizing the disutility associated to the promotion of worker B. This means that only workers B face the risk of being discriminated against, because of the observable characteristic that, without affecting their productivity, differentiates them from workers A. Since productivity is assumed to be entirely paid to workers, in this model discrimination can only assume the form of denying a promotion to a worker B that would deserve it. Employer's utility function is

$$U^{\theta_F} = (m-1)(\pi_A^1 + \pi_B^1 + \pi_A^2 + \pi_B^2) - I(\alpha_B)\theta_F$$

where $m > 1$ is a known and constant mark up on workers' productivity justified by the entrepreneurial activity. Therefore, in order to maximize profits, the employer needs to maximize worker's productivity and has therefore the incentive to promote the more productive worker.[11] The term $I(\alpha_B)\theta_F$ represents the disutility associated to a promotion of a minority worker. When $\theta_F = 0$ the observable characteristic that distinguishes the workers does not matter and profits are the only thing that the employer considers. On the other hand, when $\theta_F = d$ the employer is characterized by discriminatory tastes.

3.1.3 Strategies

Workers simultaneously choose effort twice, the second time after the decision about promotion has been taken by the employer. The strategy s of a worker is therefore a triple containing an effort level for the first period, and two effort levels for the second period, one if promoted, another if not promoted. The optimal choice in the second period has already been computed above. The optimal effort in the first period, derived taking first order conditions of equation 2, is instead:

$$e_P^{1*} = 1 + \frac{\partial \mu_P(\alpha_P(e_P^1))}{\partial e_P^1}(\theta_P - 1), \tag{3}$$

The interpretation is straightforward. Optimal effort in the first period depends on its impact on the probability of being promoted, and such an impact is weighted according to the worker's ability. Note that when there is no chance of being promoted, the last term of these equation vanishes and $e^1 = 1$ becomes the optimal choice. In other words, in a world without promotions or, alternatively, when all the employers are discriminatory, only the instantaneous utility function in the first period matters.

[11] It is also possible to interpret F as a supervisor rather than an employer. Instead of profits, the supervisor maximizes a bonus which is a proportional to the overall productivity of the supervisees. Nothing changes, because also the supervisor has the incentive to promote the more productive supervisee in order to maximize his/her bonus.

The employer observes each worker's productivity in the first period and then promotes one (and only one) of them in a more rewarding position. The set of feasible actions for the employer, regardless of his/her type, is simply $S_F = \{\alpha_A; \alpha_B\}$. As far as the employer is concerned, strategies s^F are therefore given by a function that specifies a promotion decision for every possible pair of observed productivity levels.

To complete the description of the constituent game, also players' beliefs need to be specified. To do this, however, it is necessary to describe how players are matched and what information they can access.

3.2 The Population Game

The constituent game described in section 3.1 is inserted in a wider game, called population game, which specifies how players are matched and what information they can access. The description of the information structure allows to define players' beliefs.

There are three populations, one of employers and two of workers. As already said for the constituent game, the two populations of workers differ because of an observable characteristic (gender, race, etc.) that does not affect their productivity. The distribution of types within the two populations of workers is identical. This assumption rules out the possibility that unbalanced promotions across populations arise because of a different average ability.

3.2.1 Matching

Each of the three populations $P = \{A, B, F\}$ is composed by a continuum of players. At every stage each player of population A is randomly matched with one player from population B and one player from population F.

3.2.2 Information Structure

Players try to figure out the distribution of types and strategies among the populations of opponents using available information. Besides observing the sequence of actions $\sigma = (e_A^1, e_B^1, \alpha, e_A^2, e_B^2) \in \Omega$ of the constituent game where she is involved, every player also observes the distribution of promotions between populations $\hat{\alpha}$. At first sight individual information seems to be more informative, since it allows players to compute the probability of being promoted conditional on workers' output, while aggregate information does not. However, if every worker participates in the labor market for a small number of rounds, as it seems plausible to assume, individual information becomes negligible as compared to aggregate information.

3.2.3 Beliefs

Beliefs of a player are a probability measure over the unknown component of the belief-type-strategy set $\Theta \times M \times S = \Theta_A \times \Theta_B \times \Theta_F \times M_A \times M_B \times M_F \times S_A \times$

$S_B \times S_F$. This implies that beliefs are not assumed to be common knowledge and therefore players must hold second order beliefs about opponents' expectations.

Given that every player is supposed to know her type, her beliefs and the strategy she chooses only, the unknown component of $\Theta \times M \times S$ turns out to be the set of belief-type-strategy profiles of all the other players, both the opponents and the other players of his/her own population. Beliefs are assumed to be equal within each population, i.e. to by independent from the type and beliefs of the players. Beliefs of a worker of population P are therefore defined $\mu_P \in \Delta(\Theta \times M \times S)$. In what follows I will refer to the "type" of a worker meaning only her ability, and to the "belief-type" meaning the combination of her ability and her beliefs.

Beliefs are *correct* whenever, for all the belief-types $\theta \times \mu$ of each population, the subjective probability distribution over opponents' belief-type-strategy set coincides with the objective distribution. For instance,

$$\mu_A(\theta_B, \mu_A, s_B, \theta_F, \mu_F, s_F) = \Pr(\theta_B, \mu_B, s_B, \theta_F, \mu_F, s_F) \tag{4}$$

intuitively means that the probability assigned by each (belief-type of) player of population A to every combination of opponents' belief-type-strategy profile is correct.

Beliefs are *not contradicted by the evidence* whenever all the players' subjective probability to observe a particular distribution of aggregate outcomes $\hat{\alpha}$ coincide with its actual frequency. The subjective probability of observing $\hat{\alpha}$ is obtained summing up the probabilities attached to every combination of opponents' belief-type-strategy profiles that leads to a combination of observables equal to $\hat{\alpha}$. It may happen that incorrect beliefs, i.e. beliefs which violate (4) for some belief-type $\theta \times \mu$ or strategy s of the opponents, are not contradicted by the evidence.

The intermediate case, when beliefs are correct only as far as the behavioral rules are concerned, can be represented in the following way:

$$\begin{array}{rcl} \mu_A(s_B, s_F|\theta_B, \mu_B, \theta_F, \mu_F) & = & \Pr(s_B, s_F|\theta_B, \mu_F, \theta_F, \mu_F) \\ \mu_A(\theta_B, \mu_B, \theta_F, \mu_F) & \neq & \Pr(\theta_B, \mu_B, \theta_F, \mu_F). \end{array}$$

This case is particularly important, because Bayesian equilibria require players to hold correct beliefs concerning the behavioral rules of the opponent, while beliefs about belief-types may be wrong. The equilibria driven by wrong expectations of being discriminated against are characterized by these features, and in addition they are not contradicted by the evidence (see below, Section 4 Proposition 2).

What follows is a list of assumptions with the goal of neutralizing as much as possible all the sources of unequal outcomes other than the role of workers' expectations. Although admittedly tedious, I believe that such a list is useful because it has the advantage of making explicit the information structure within which workers' expectations play a role.

Assumption 2.

a) *Belief-type-strategy profiles of opponents are not correlated.* Since every player knows her own belief-type and strategy, the appropriate marginal distribution for worker A is:

$$\mu_A(\theta_B, \mu_B, s_B, \theta_F, \mu_F, s_F) = \mu_A(\theta_B, \mu_B, s_B)\mu_A(\theta_F, \mu_F, s_F).$$

b)*Players hold correct expectations about behavioral rules of each belief-type of opponent.* This makes unnecessary to specify beliefs about opponents' strategies. In fact, this assumption means that once a belief-type is identified, expectations about her strategy are correct:

$$\mu_A(\theta_B, \mu_B, s_B)\mu_A(\theta_F, \mu_F, s_F) = \mu_A(\theta_B, \mu_B)\mu_A(\theta_F, \mu_F).$$

c)*Beliefs are type-independent.* Meaning that beliefs are the same for all players of the same population, this finally allows to specify players' beliefs in the most parsimonious way, to which I will sometimes refer simply as μ_A:

$$\mu_A(\theta_B, \mu_B)\mu_A(\theta_F, \mu_F) = \mu_A(\theta_B)\mu_A(\mu_B)\mu_A(\theta_F)\mu_A(\mu_F).$$

Something more should be said about employers' beliefs. In fact, before deciding, the employers have the opportunity to update their beliefs observing workers' productivity. $\mu_F(\theta_A, \theta_B | e_A^1, e_B^1)$ are the updated beliefs of an employer about the type of workers having observed the pair of effort levels $e_A^1, e_B^1)$ in the first.

Assumption 3. *Workers believe that the strategies of the employers are weakly monotone in effort:*

$$\begin{array}{rcll}\mu_A\left(\alpha_A | e_A^1, e_B^1\right) & \geq & \mu_A\left(\alpha_A | \hat{e}_A^1, e_B^1\right), & \quad \forall e_A^1 > \hat{e}_A^1, \forall \hat{e}_A^1, e_B^1, \\ \mu_B\left(\alpha_B | e_B^1, e_A^1\right) & \geq & \mu_B\left(\alpha_B | \hat{e}_B^1, e_A^1\right), & \quad \forall e_B^1 > \hat{e}_B^1, \forall \hat{e}_B^1, e_A^1.\end{array}$$

Workers think that the probability of being promoted cannot decrease when effort increases *ceteris paribus.* This assumption is a way to refine the set of equilibria. The intuitive reason is that promotions are desirable, and workers are willing to give up some utility in the first period in order to enhance their probability of being promoted. The way to do this is to deviate from $e_1 = 1$ in the first period supplying either a higher or a lower effort. Assumption 3 implies that the derivative in equations 3 is positive, and therefore one's willingness of being promoted must be signaled only through a higher effort. In this way, all the equilibria that could possibly arise when workers signal their willingness of being promoted supplying a lower effort are excluded, because exerting an effort lower than 1 in the first period turns out to be strictly dominated for all workers.[12]

[12]The game can be considered like an all-pay auction, insofar as the utility loss suffered by a non-promoted worker who chooses effort $e^1 > 1$ is sunk, see Baye, Kovenock and de Vries (1996).

Although this paper does not explicitly deal with dynamics, I think it is useful to provide an intuition about how beliefs may be formed. Beliefs of a player at time t can be thought to be a function of the available information about aggregate outcomes (promotions) arising from the previous period $\hat{\alpha}_{t-1}$. Since the paper focuses on the equilibria of the game, time subscripts will be omitted in what follows. Notice that the same sequence of observables can lead to different beliefs. In other words, players can interpret in different ways the same information about promotions. For example, workers can interpret a given distribution of promotions across populations A and B assigning different weights to the role played by workers' ability as opposed to employers' propensity to discriminate against the minority. Of course, asymptotic empiricism requires that in equilibrium all the belief rules must generate subjective distributions of observables which coincide with the objective one.

Assumption 4. *Beliefs about the belief-type-strategy profile of workers A*: only minority workers may have wrong beliefs about the type-strategy profile of workers A (a). Although their expectations about the distribution of types within majority workers are correct (b), they may hold wrong beliefs about their expectations (c). The reason is that minority workers are assumed to believe that the two populations of workers are identical also in terms of beliefs.

a) $\mu_A(\theta_A, \mu_A) = \mu_F(\theta_A, \mu_A) = \Pr(\theta_A, \mu_A)$

b) $\mu_B(\theta_A) = Pr(\theta_A)$

c) $\mu_B(\mu_A) = \mu_B$.

In particular, if minority workers think there are discriminatory employers, as explained in the next assumption, they think that also workers A share the same beliefs and behave accordingly.

Assumption 5. *Beliefs about employers' belief-type-strategy profile*: only minority workers may have wrong beliefs about the distribution of employers' profiles (a). This captures their possible expectations of being discriminated against as induced by observed unequal outcomes (b). Minority workers think that employers share the same beliefs as them (c).

a) $\mu_F(\theta_F, \mu_F) = \mu_A(\theta_F, \mu_F) = \Pr(\theta_F, \mu_F)$

b) $\mu_B(\theta_d) = 2(0.5 - \hat{\alpha}_B)$

c) $\mu_B(\mu_F) = \mu_B$.

The belief rule of workers B implies that they attribute the entire gap in the distribution of promotions to employers' discriminatory tastes. In fact, only when promotions are balanced they think that all the employers are unbiased. Since they believe that the two populations are identical (see Assumptions 4a, 4c and 6a), if the fraction of workers B promoted is less than 0.5 their belief rule attributes the whole difference to the presence of a corresponding fraction of discriminatory employers. It is worth stressing that Assumption 5 implies a possible violation of the common prior assumption given that beliefs about employers' type may differ between workers A and B.

Assumption 6. *Beliefs about the belief-type-strategy profile of workers B*: only employers may have wrong beliefs about the profile of workers B (a). This captures a sort of statistical discrimination effect that rationalize the possibly unbalanced distribution of promotions (b) based on their beliefs that workers B share the same expectations as them (c). What happens is that employers, thinking that there are no discriminatory employers (see Assumption 5a) believe that workers B think the same and behave accordingly. Therefore, a way to rationalize possibly unbalanced promotions is a belief rule that attributes to minority workers' ability what in fact is the interaction between ability and expectations of being discriminated against.

a) $\mu_B(\theta_B, \mu_B) = \mu_A(\theta_B, \mu_B) = \Pr(\theta_B, \mu_B)$

b) $\mu_F(\theta_B) = U\left[1, 1 + 2\hat{\alpha}_B\right]$

c) $\mu_F(\mu_B) = \mu_F$.

It is worth stressing that wrong beliefs of employers concerning the distribution of ability within minority workers have no behavioral implications. On the contrary, they are simply a way to rationalize the possibly unbalanced distribution of promotions. The first period of the game can be regarded as a trial work period, after which it is still optimal to promote the worker displaying the higher effort (see below, section 4). Such an assumption is necessary in this case, because had they correct beliefs about minority workers' expectations, they would manage to invert their choice function (see equation 7 below) and uniquely determine workers' type. This would trivially imply that employers would have complete information about workers' type at the end of the first period, while, in line with an old saying, some uncertainty is necessary for a horse race to take place.

To summarize, expectations are assumed to be correct, apart from:

a) minority workers who may (wrongly) think there is a positive fraction of discriminatory employers, and in this case they think that all players share the same beliefs as theirs;

b) employers who rationalize possibly unbalanced promotions revising (wrongly) their beliefs about minority workers' ability, since they believe that there are no discriminatory employers and they think that all players share the same beliefs as theirs.

Two things needs to be stressed. First, among the possibly wrong beliefs, only minority workers' expectations of being discriminated against have behavioral implications. In fact, as it is shown below (see Proposition 2), the actual fraction of workers B promoted depends only on $\mu(\theta_0)$. This should not be surprising that most of the assumptions made so far are aimed at isolating the role of workers' expectations. It deserves to be stressed once more that such assumptions are made with the only purpose to focus the theoretical analysis on the role of workers' expectations and not because the other causes of unequal outcomes are regarded as less important. Second, it may seem weird that, according to Assumptions 4-6 above, all players attribute to the opponents the same beliefs they hold, apart from workers A who instead are characterized by correct second order beliefs. In other words, workers A think there are no discriminatory employers but correctly believe that minority workers think that such a fraction is positive. This is the most parsimonious way to get the results in Section 4, but equivalent results could be obtained even assuming that workers A attribute to workers B correct beliefs that there are no discriminatory employers, provided that at the same time majority workers are characterized by self-serving beliefs in order to rationalize possibly unbalanced promotions (see section 4.3).

4 Analysis of the equilibria

Two different concepts are necessary to analyze the equilibria of the model: the Perfect Bayesian Equilibrium (henceforth: PBE) and the Self-Confirming (or Conjectural) Equilibrium (henceforth: SCE).[13] The two concepts share the feature that each player maximizes utility given her beliefs, updated whenever possible, about every possible opponents' profile (see section 4.1). The difference between them is that in a PBE each player has a correct conjecture about the relationship between opponents' belief-types and choices, i.e. about their behavioral rules. In the commonly applied subcase when beliefs satisfy the Common Prior assumption, beliefs about opponents' types are correct as well. On the other hand, when the Common Prior assumption is not satisfied, beliefs in a PBE may be contradicted by the evidence. On the contrary, in a SCE beliefs may not coincide with the true distribution of opponents' belief-types and expectations about opponents' behavioral rules may also be wrong, as long as they are not contradicted by the evidence (see Battigalli and Guaitoli (1997) and Dekel, Fudenberg and Levine (2002) for a formal discussion of the relation between the Common Prior assumption, PBE and SCE in games of incomplete information).

In this section two qualitatively different sets of equilibria are presented. The first set (see Proposition 1) displays symmetric outcomes under the assumption that minority workers hold correct expectations that there are no discriminatory employers. The second set of equilibria (see Proposition 2) shows that unequal outcomes arise when minority workers' expectations are wrong *ceteris paribus*. All these equilibria are Perfect Bayesian and Self-Confirming at the same time. In fact, players always predict correctly the opponents' behavioral rules. Moreover, beliefs are never contradicted by the evidence. This is fairly intuitive in Proposition 1 given that the Common Prior assumption is satisfied and therefore beliefs turn out to be correct. On the other hand, this is also true considering the equilibria in Proposition 2, although they do not predict correctly the belief-type of the opponents in some cases.

4.1 Utility maximization given beliefs

a) Employers

[13] For a thorough exposition of the characteristics of SCE the reader is referred to Battigalli (1987) and Fudenberg and Levine (1993 and 1998). The generalization of the SCE to the case of aggregate outcomes is described in Filippin (2003a).

Only workers' difference in productivity *after* the promotion affects employer's decision, while the difference in the first period does not. The reason is that the disutility θ_F is associated to the *promotion* of a minority worker. Therefore, at the margin only benefits from the promotion of a minority worker (i.e. difference in productivity after promotion) are compared with the associated cost θ_F in order to decide which worker is optimal to promote.

Employers of type θ_d are characterized by tastes for discrimination that are assumed to be so high that they promote worker A regardless of any observed and expected productivity level of the two workers:

$$\alpha_A = BR^{\theta_d}(\pi^1).$$

Employers of type θ_0 are instead unaffected by the observable characteristic that distinguishes workers A from workers B, and therefore they do not suffer a disutility promoting a minority worker. Hence, they will always promote the worker they think will be more productive *after* the promotion, i.e. the worker characterized by higher ability, regardless of the population where he/she comes from.

Defining $\mu^{\theta_0}(\pi_A^2|\pi_A^1)$ the updated beliefs of a non-discriminatory employer about the productivity π^2 of worker A in the second period having observed π_A^1 in the first, it follows that the best reply $BR^{\theta_0}\left(\pi^1|\mu^{\theta_0}\right)$ to the observed pair of productivity levels $\pi^1 = (\pi_A^1, \pi_B^1)$ will depend on the comparison of workers' expected productivity in the second period. Exploiting the assumption that $\pi^1 = e^1$ and since employers correctly believe that productivity of a promoted worker is strictly increasing in her ability, we get that:

$$\begin{gathered}\alpha_A \in BR^{\theta_0}\left(e^1|\mu^{\theta_0}\right) \text{ if} \\ \int_1^2 \mu^{\theta_0}(\theta_A|e_A^1)\theta_A \; d\theta_A > \int_1^2 \mu^{\theta_0}(\theta_B|e_B^1)\theta_B \; d\theta_B \end{gathered} \tag{5}$$

which means that promoting a majority worker is the best reply whenever the expected ability of the majority worker is higher, given the observed productivity levels. Similarly, promoting a minority worker is the best reply whenever equation 5 holds with reversed inequality sign. If expected productivity in the second period is the same, the non-discriminatory employer is indifferent. This means that both α_A and α_B as well as all the mixed strategies would be best replies. Employers believe that all workers think there are no discriminatory tastes (see Assumptions 4a, 5a and 6b), and therefore they update their beliefs about workers' type inverting the same effort function for both populations (equation 6 below). Hence, given the strictly monotone relation between effort and ability, the best reply of an unbiased employer is to implement a fair tournament and to promote the worker who displays a higher effort.

b) Workers

Assumption 5 implies that workers correctly guess that an unbiased employer implements a fair tournament. This is true for workers A, who have correct beliefs, but also for workers B, because they think that the two populations of workers are equal (also in terms of beliefs) and they attribute the same beliefs to the employers. Therefore, they correctly anticipate that unbiased employers maximizes profits promoting the worker displaying the higher effort, because it implies a higher ability. On the other hand, everybody knows that discriminatory employers always promote worker A. Therefore, expectations of being promoted for a minority worker are equal to the (subjective, and possibly wrong) probability of facing an unbiased employer times the probability that his/her effort is greater than that of the opponent.

$$\mu_B(\alpha_B(e^1)) = \mu(\theta_0)\mu(e^1_B > e^1_A).$$

Majority workers are also assumed to have correct beliefs about the percentage of discriminatory employers, $\mu_A(\theta_0) = Pr(\theta_0) = 1$, and therefore:

$$\mu(\alpha_A(e^1)) = \mu(e^1_A > e^1_B).$$

Substituting these results into equation 3, which displays the optimal choice of effort in the first period, we obtain:

$$e^{1*}_A = 1 + \frac{\partial \int_1^{e^1_A} \mu(f(e^1_B))de^1_A}{\partial e^1_A}(\theta_A - 1),$$
$$e^{1*}_B = 1 + \mu(\theta_0)\frac{\partial \int_1^{e^1_B} \mu(f(e^1_A))de^1_B}{\partial e^1_B}(\theta_B - 1),$$

where $\mu(f(e))$ is the expected distribution of effort in the population of opponents. Note that incentive to exert an effort greater than one depends on the *change* of the probability of being promoted due to a higher effort and not on its *level*. Therefore, even if there are discriminatory employers, incentives are the same for both populations, because the assumption that discriminatory employers always promote A makes the incentive to exert effort proportional to the percentage of non-discriminatory employers. In the limit situation where there are only discriminatory employers, promotions are no more an incentive device for both populations, because A are sure of being promoted, while B have no chance. This parallels the finding within unfair tournaments that both agents exert the same level of effort in equilibrium. Of course, the presence of a strictly positive fraction of non-discriminatory employers is necessary for promotions to work as an incentive device.

Supposing that workers believe that among opponents effort is uniformly distributed with density u, we finally get:

$$e^{1*}_A = 1 + \mu_A(u_B)(\theta_A - 1), \tag{6}$$
$$e^{1*}_B = 1 + \mu_B(\theta_0)\mu_B(u_A)(\theta_B - 1), \tag{7}$$

with effort that is a linear function of ability, and therefore actually distributed in a uniform way, with the upper bounds depending on expectations about the fraction of discriminatory employers and the lower bound equal to one in both populations, corresponding to the worker with the lowest ability $\theta = 1$.

4.2 The equilibria of the game

Let us see first what happens when minority workers have correct expectations $\mu_B(\theta_d) = \Pr(\theta_d) = 0$ that there are no discriminatory employers (Proposition 1). What found in this way is then used as a benchmark in the other case, $\mu_B(\theta_d) > 0$, i.e. when minority workers overestimate the fraction of discriminatory employers (Proposition 2). Considering the assumptions made so far, it is worth stressing that minority workers' expectations about employers' type are the only variable with behavioral implications, and therefore the only source of unequal outcomes. The other beliefs that may be wrong, namely statistical discrimination of employers, are instead just a way to rationalize unequal outcomes, but they do not affect players' behavior.

Proposition 1 *When expectations of workers B about employers' type are correct a Perfect Bayesian and Self-Confirming Equilibrium always exists, characterized by:*

1) every worker B that chooses in the first period the same effort of the corresponding type of population A.

2) balanced promotions across populations.

Sketch of proof: When $\mu_B(\theta_0) = 1$, the choice of optimal effort becomes the same for majority and minority workers (see equations 6 and 7). From

$$u_A = u_B = u^* = \frac{1}{max(e^*) - min(e^*)}$$

it is possible to derive that $u^* = 1$ and therefore that effort is uniformly distributed between 1 and 2 in both populations. Therefore, due to random matching, half of the times effort of worker B will be higher than effort of worker A. Given the optimal strategy of the employers, all unbiased by assumption, it turns out that promotions are balanced. All beliefs are trivially correct, given that all the possible sources of mistakes vanishes when promotions are balanced.

Proposition 2 summarizes what changes if minority workers' expectations are instead wrong, and in particular when they overestimate the percentage of discriminatory employers.

Proposition 2 *When expectations of workers B overestimate the percentage of discriminatory employers ($\mu(\theta_0) < 1$) a Perfect Bayesian and Self-Confirming Equilibrium always exists, characterized by:*

1) every worker B in the first period that chooses a lower effort than the corresponding type of population A.

2) unbalanced promotions: $\hat{\alpha}_B = 0.5\mu_B(\theta_0)$.

Sketch of proof: In this case the expected return to the effort exerted by minority workers is weighted down by the factor $\mu_B(\theta_0) < 1$ as compared to that of majority workers *ceteris paribus* (see equations 6 and 7). Following

Assumption 4c, workers B think that workers A behave in the same way as them, and therefore imposing $\mu_B(u_A) = u_B = u^*$ and solving

$$u^* = \frac{1}{max(e_B^*) - min(e_B^*)} = \frac{1}{(e_B^*|\theta = 2) - (e_B^*|\theta = 1)} = \frac{1}{\mu_B(\theta_0)u^*}$$

we obtain that

$$u^* = \frac{1}{\sqrt{\mu_B(\theta_0)}},$$

which substituted into equation 7 finally leads to

$$e_B^* = 1 + \sqrt{\mu_B(\theta_0)}(\theta_B - 1).$$

Effort in population B is therefore uniformly distributed between 1 and $\sqrt{\mu_B(\theta_0)}$.

Workers A are assumed to have correct beliefs about the profile of workers B. Therefore, they correctly anticipate that

$$\mu_A(u_B) = u_B = \frac{1}{\sqrt{\mu_B(\theta_0)}},$$

which substituted into equation 6 leads to

$$e_A^* = 1 + \frac{1}{\sqrt{\mu_B(\theta_0)}}(\theta_A - 1).$$

However, it is pointless for workers A to exert an effort greater than the maximum of the opponents, because it would imply a strictly lower utility in the first period without increasing the probability of being promoted which is already equal to one, and therefore a strictly dominated strategy. Therefore effort in population A will be uniformly distributed between 1 and $\sqrt{\mu_B(\theta_0)}$, with density $u_A = \sqrt{\mu_B(\theta_0)}$ and there will be a mass of probability equal to $1 - \mu_B(\theta_0)$ at the effort level $e_A^{1*} = 1 + \sqrt{\mu_B(\theta_0)}$. This does not affect the choice of workers B, given their expectations, and therefore nothing changes in terms of promotions. Notice that expectations of being discriminated against create a stronger incentive to exert a higher effort for the workers A. In fact, effort of workers B belongs to a smaller interval, and the higher density increases at the margin the gain that workers A derive from a higher effort in terms of probability of being promoted.

The optimal strategy of the employers implies that workers A will always be promoted whenever

$$e_A^1 > max(e_B^1) = \sqrt{\mu_B(\theta_0)},$$

and given that the density $f(e_A^1)$ is $u_A = \sqrt{\mu_B(\theta_0)}$, this implies that the best

$$(\frac{1}{\sqrt{\mu_B(\theta_0)}} - \sqrt{\mu_B(\theta_0)})\sqrt{\mu_B(\theta_0)}) = 1 - \mu_B(\theta_0)$$

percent of workers A are certainly promoted, together with half of the other $\mu_B(\theta_0)$ in which effort of the majority worker is higher. Similarly, none of the $1-\mu_B(\theta_{=}0)$ workers B matched with the best workers A are promoted, while only half of the other $\mu_B(\theta_0)$ are promoted. In this game promotions, and in particular whether they are balanced or not, depend only on minority workers' expectations of being discriminated against.

As already mentioned, the equilibria in Proposition 1 and Proposition 2 are not only Perfect Bayesian but also Self-Confirming equilibria. Having seen that every player is maximizing her utility given beliefs which are correct concerning opponents' behavioral rules is enough to ensure that the equilibria are PBE. On the contrary, to show that they also are SCE it is also necessary to check that beliefs are not contradicted by the evidence.

Empiricism

Workers A have correct beliefs about other players' profiles. Hence, the objective distribution of observables must coincide with the subjective distributions implied by their beliefs.

Workers B will not update their wrong beliefs about the fraction of discriminatory employers $\mu_B(\theta_d)$. In fact, substituting the fraction of minority workers actually promoted $0.5\mu_B(\theta_0)$ into their belief rule described in Assumption 5b we obtain a fixed point:

$$\mu_B(\theta_d) = 2(0.5-\hat{\alpha}_B) = 1-\mu_B(\theta_0) = \mu_B(\theta_d).$$

Given that beliefs of minority workers do not change, even the percentage of workers B promoted do not change. As a consequence, beliefs of employers will not be updated as long as they are equal to:

$$\mu_F(\theta_B) = U\left[1, 1+2\hat{\alpha}_B\right].$$

4.3 Self-serving beliefs

As mentioned at the end of Section 3.2.3, the same results of Proposition 2 also be obtained assuming that workers A think that the other players share the same beliefs as them. Namely, this would imply that, instead of holding correct second order beliefs concerning minority workers' expectations of being discriminated against, workers A believe that also minority workers correctly guess that there are no discriminatory employers. However, to obtain the same results, majority workers need to rationalize unbalanced promotions that would be observed, given that workers B would exert a lower effort than that predicted by workers A. A way to do this is to assume that majority workers are characterized by self-serving beliefs, taking the form of an overestimate of their individual as well as population ability.

In more detail, workers A expect effort of the opponents to be uniformly distributed between 1 and 2, while indeed it belongs to the interval $\left[1, \sqrt{\mu_B(\theta_0)}\right]$. Therefore, unbalanced promotions would unexpectedly emerge according to workers A subjective distribution of observables. According to the evidence reported by Filippin and Ichino (2005), groups that experience favorable outcomes are likely to interpret such outcomes in a way that enhances their self-esteem. For instance workers A, while keeping an objective view of the opponents' ability, could overestimate their own ability.

If their own perceived ability is

$$\overline{\theta}_A = \frac{\theta_A}{\sqrt{2\hat{\alpha}_B)}}$$

the distribution of effort within population A becomes the same as in Proposition 2 and therefore also the actual fraction of minority workers promoted becomes the same $(0.5)\mu_B(\theta_0)$. However, such fraction would be consistent with beliefs of minority workers and employers, but not with those of workers A, who would expect a higher fraction, namely $\sqrt{2\hat{\alpha}_B}$, of workers B promoted. For the outcomes to identify a Self-Confirming Equilibrium it is necessary that, in addition, workers A overestimate group ability. In particular it must be that:

$$\mu_A(\theta_A) = U\left[1, \frac{1}{\hat{\alpha}_B}\right]$$

4.4 Policy implications

Trial work periods can be an effective policy tool to break down statistical discrimination outcomes, i.e. a situation where employers' wrong beliefs are self-confirming. On the contrary, the equilibria described in Proposition 2 are robust to trial work periods, for the simple reason that trial work periods affect employers' expectations rather than workers' expectations. As long as minority workers think of being discriminated against, during the first period that can be regarded as a long trial work period they display on average a lower productivity. At the end of the first period even unbiased employers are therefore more likely to promote workers A.

Quotas can also be implemented to correct unequal outcomes. However, despite being effective to increase the number of minority workers promoted, quotas do this without affecting the mechanism that generates unequal outcomes in equilibrium. Therefore, the effect of quotas is predicted to be transitory.

The simplest way to implement quotas is to impose that at least a percentage $q > 0$ of minority workers must be promoted, with q known by all players. In this model, given that only one worker from each population participates to every constituent game, such a result can be obtained imposing a lottery to the employers. The outcomes of this lottery are that with probability q employers are forced to promote the minority worker, while with probability $1 - q$ they are free to choose according to their preferences and updated beliefs. Expected utility becomes therefore

$$
\begin{aligned}
EU^{\theta_A}(e_A^1) &= e_A^1 - \frac{(e_A^1)^2}{2} + 1 + (1-q)\mu(\alpha_A(e_A^1))(\theta_A - 1), \\
EU^{\theta_B}(e_B^1) &= e_B^1 - \frac{(e_B^1)^2}{2} + 1 + q(\theta_B - 1) + (1-q)\mu(\alpha_B(e_B^1))(\theta_B - 1).
\end{aligned}
$$

Paradoxically, after the introduction of quotas, workers exert an even lower effort as compared to equations 6 and 7:

$$
\begin{aligned}
e_A^{1*} &= 1 + (1-q)\mu_A(u_B)(\theta_A - 1), \\
e_B^{1*} &= 1 + (1-q)\mu_B(\theta_0)\mu_B(u_A)(\theta_B - 1).
\end{aligned}
$$

The fraction of workers B promoted will increase from $0.5\mu(\theta_0)$ to $q + 0.5(1-q)\mu(\theta_0)$, but without affecting minority workers' expectations. Once quotas are removed, the fraction of minority workers promoted should go back to the old level $0.5\mu(\theta_0)$. Quotas could be effective in a more general version of the model, when more than one worker from each population participates in the same tournament, provided that the competition induced between minority workers is strong enough to make them exert a sufficiently high effort. If this is the case, the rate of promotions within minority workers increases by more than what directly induced by quotas, imposing the minority workers to update their beliefs toward a lower expected fraction of discriminatory employers.

Dealing with feedback effects model, Cain (1986) raises a concern which also applies to this model and, more generally, to all the models displaying multiple equilibria some of which suboptimal:

> "model's predicted consequences from a favorable shock are so obviously beneficial to the group discriminated against and to employers that is difficult to see why the upward spiral would not quickly be initiated by group intervention."

This argument suggests that it should not be difficult to break down unequal outcomes based on workers' expectations, and this is certainly true as far as the mathematics of the model is concerned. Many devices can perform this task, like a subsidy to minority workers proportional to their effort, or a free insurance that pays back the money equivalent of the utility loss suffered by minority workers who supplied a high effort without being promoted, and so on. Nevertheless, this devices do not seem to have an intuitive counterpart on the field, due for instance to the impossibility of enforcing contracts based on unobservable effort. The bottom line is that, in line with Coate and Loury (1993), the best way to correct unequal outcomes is to affect expectations of minorities.[14] Policy tools which do not change the expectations of minorities are either ineffective or very difficult to implement.

[14] For instance, the Gay Pride can also be thought as a device that reduces the sexual minorities' expectations to be discriminated against in the labor market.

5 Conclusions

The model presented in this paper analyzes from the theoretical point of view the role of workers' expectations, so far neglected in the literature, in explaining the observed unequal outcomes that characterize some minorities in the labor market.

Workers' expectations may differ even though players are given access to the same information. In other words, different beliefs are not justified by different learning processes. On the contrary, workers interpret the same aggregate observables in a different, and self-serving, way.

In line with the evidence reported by Filippin and Ichino (2005), unequal outcomes are rationalized by minority workers via expectations of being discriminated against. The same unequal outcomes may instead be rationalized by majority workers by means of self-serving beliefs concerning the distribution of ability across populations, i.e. overestimating their individual as well as population ability in a way that enhances their self-esteem.

The Common Prior hypothesis is therefore violated. As a consequence, beliefs in a Perfect Bayesian Equilibrium may be contradicted by the evidence, in such a way that this commonly used equilibrium concept could not be associated with a fixed point of a learning process. For this reason, the equilibria presented in this paper are also self-confirming, i.e. not contradicted by the evidence, so that they can safely be interpreted as a fixed point of a learning process

The importance of workers' expectations can be appreciated comparing the equilibrium outcome in terms of promotions arising when minority workers overestimate the percentage of discriminatory employers as compared to a situation in which such beliefs are correct *ceteris paribus*. Even in a labor market where employers do not discriminate against minority workers, and where the distribution of ability is the same across groups of workers, unequal outcomes may still arise due only to workers' expectations. Unequal outcomes are rationalized by the employers using a statistical discrimination argument, which however does not have behavioral implications and therefore does not affect the distribution of promotions. On the contrary, unbiased employers implement a fair tournament.

The result that that unequal outcomes can be ascribed to workers' different expectations, is robust both to trial work periods, which are instead an effective policy device to break down statistical discrimination outcomes, and to affirmative actions like quotas.

The conclusion is that workers' expectations can contribute to explain why historically oppressed social groups are not likely to forge ahead once the original cause of unequal outcomes has been removed, confirming the evidence emerging from laboratory and field experiments.

REFERENCES

Aigner, D.J. and G.G. Cain (1977): "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30:2, 175-87.

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

Battigalli, P. (1987): *Comportamento Razionale ed Equilibrio nei Giochi e nelle Situazioni Sociali,* unpublished dissertation, Universita' Bocconi, Milano.

Battigalli, P. and D. Guaitoli (1997): "Conjectural Equilibria and Rationalizability in a Game with Incomplete Information," in *Decisions, Games and Markets* edited by P. Battigalli, A. Montesano and F. Panunzi. Dordecht: Kluwer Academic Publishers, 97-124.

Baye, M.R., D. Kovenock and C.G. De Vriers (1996): "The All-Pay Auction with Complete Information," *Economic Theory*, 8, 291-305.

Becker, G.S. (1957): *The Economics of Discrimination.* Chicago: University of Chicago Press.

——— (1985): "Human Capital, Effort and the Sexual Division of Labor," *Journal of Labor Economics,* 3, S33-S58.

Blau, F.D., M.A. Ferber and A. E. Winkler (2002): *The economics of Women, Men and Work.* Upper Saddle River, NJ: Prentice Hall.

Blau, F.D. and C.L. Jusenius (1976): "Economists' Approaches to Sex Segregation in the Labor Market: An Appraisal," in *Women and the Workplace* ed. by M. Blaxall and B. Reagan. Chicago: University of Chicago Press, 181-99.

Breen R., and C. Garcia-Penalosa (2002): "Bayesian Learning and Gender Segregation," *Journal of Labor Economics, 20:4, 899-922.*

Cain, G.G. (1986): "The Economic Analysis of Labor Market Discrimination: a Survey," in *Handbook of Labor Economics,* ed. by O. Ashenfelter and R. Layard. Amsterdam: North Holland, 693-785.

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

Corns, A. and A. Schotter (1999): "Can Affirmative Action be Cost Effective? An Experimental Examination of Price-Preference Auctions," *American Economic Review, 89, 291-305.*

Dekel, E., D. Fudenberg and D.K. Levine (2002): "Learning to Play Bayesian Games," mimeo.

Ferber, M.A. and H.M. Lowry (1976): "The Sex Differential in Earnings: A Reappraisal," *Industrial and Labor Relations Review*, 29, 377-87.

FILIPPIN, A. (2003a): "A Generalization of the Self-Confirming Equilibrium Concept to Aggregate Observables," mimeo.

——— (2003b): "Discrimination and Workers' Expectations: Experimental Evidence," IZA WP 824.

——— (2004): "The Persistence of Discrimination: a View from the Lab," mimeo.

FILIPPIN, A. AND A. ICHINO(2005): "Gender Wage Gap in Expectations and Realizations," *Labour Economics*, 12, 125-145.

FUDENBERG, D. AND D.K. LEVINE (1993): "Self-Confirming Equilibrium," *Econometrica, 61:3, 523-545.*

——— (1998): *The Theory of Learning in Games.* Cambridge and London: MIT Press.

HOFF, K. AND P. PANDEY (2004): "Belief System and Durable Inequalities," mimeo.

LAZEAR, E. AND S. ROSEN (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy, 89,* 841-864.

MINCER, J. AND S. POLACHECK (1974): "Family investments in Human Capital: Earnings of Women," *Journal of Political Economy*, 82:2, 76-108.

MORO, A., AND P. NORMAN (2003): "A General Equilibrium Model of Statistical Discrimination," *Journal of Economic Theory, ...,...*

MYRDAL, G. (1944): *An American dilemma.* New York: Harpher & Brothers.

O'KEEFFE, M., W. K. VISCUSI, AND R. J. ZECKHAUSER (1984): "Economic Contests: Comparative Reward Schemes," *Journal of Labor Economics*, 2:1, 27-56.

PHELPS, E.S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62:4, 659-661.

SCHOTTER, A. AND K. WEIGELT (1992): "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results," *Quarterly Journal of Economics*, 107, 511-39.

SPENCE, M.A. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87:3, 355-374.