View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Negative Stereotypes and Willingness to Change Them: Testing Theories of Discrimination in South Africa

Jorge M. Agüero*
University of Wisconsin-Madison

Preliminary. Please do not cite.

April 28, 2005

Abstract

This paper proposes a new test to distinguish between the two leading theories of discrimination: preference versus information. Discrimination based on preferences occurs when people behave as if they refuse to change their stereotypes about the capabilities of discriminated individuals. Those who discriminate based on information are willing to alter their stereotypes. Using data from a quasi-experiment in South Africa, I test for discrimination against women and non-whites. The preliminary results show no discrimination against the former. In the case of racial discrimination, players' stereotypes benefit non-whites instead of white opponents, but they are reluctant to change their impression for the former. However, they are willing to change their initial impression about white opponents. This has severe implications about the permanency of affirmative action policies.

*Department of Agricultural and Applied Economics, University of Wisconsin-Madison, 427 Lorch St, Madison WI 53706; email:jmaguero@wisc.edu. I would like to thank Michael Carter and James Walker for their comments and suggestions. Chantal and Crystal Munthree and Ingrid Woolard helped with the data collection. Gregg Dardagan helped with Epi Info 6. Michele Back helped me editing this manuscript. All remaining errors are my own.

"The economy of South Africa is like a two-storey house without a connecting staircase" -Thabo Mbeki, President of South Africa[1].

1 Introduction

The persistent income inequality observed in several developing countries has led scholars to pay attention to the existence of poverty traps[2]. However, the people at the bottom storey of the house in resident Mbeki's remarks are also of a different race from those at the top, creating the need for policies to empower those suffering from discrimination. Coate and Loury (1993) show that deciding between a temporary or permanent affirmative action policy depends on how much employers' negative stereotypes about the capabilities of discriminated workers can be eroded. Hence, the effectiveness of antidiscrimination policies depends on how much people are willing to change their prior beliefs. The goal of this paper is to estimate that willingness to change.

Recent work by Bertrand and Mullainathan (2004) is an example of the use of experiments to test for discrimination[3]. In this study, the authors sent fake curriculum vitae in response to employment ads, where the resumes were exactly the same except for the name of the applicants. Some contained names associated with whites, such as Emily and Greg, and others had names mostly associated with blacks, such as Lakisha and Jamal. The authors found that "white names" received 50% more callbacks than "black names."

We can imagine the process of employers reviewing the resumes sent by Bertrand and Mullainathan (2004) as a two-step approach. First, when Lakisha's resumes arrives, a discriminating employer infers her race by reading her name and uses a negative stereotype (or prior belief) linked to her race, to guess Lakisha's productivity. In the second step, employers receive (noisy) information about Lakisha's productivity by reading her resumes. Employers who discriminate based on preferences are not influenced by the contents of the resume. They behave as if they are unwilling to learn about her productivity, and discard her resume. Other employers might change their initial belief when the contents of the resume reflect "good signals" about her productivity but the new belief might be not high enough, so her resume is still discarded. These employers are willing to learn, hence are discriminating

[1]October 9th, 2003, Black Management Forum address.

[2]See Carter and Barret (2005) and the references therein.

[3]See section two below for references of other papers in this area as well as alternative estimation methods

based on information. The reasons of discrimination are different in each case and require different policies. Therefore, knowing that Lakisha's resume was rejected is clear evidence of discrimination but does not tell us which policies would be more effective to reduce it.

To understand the reasons explaining discriminatory behavior, this paper introduces a model of learning as an analog of reviewing a resume. I model the behavior of a group of individuals who do not know each other. They have to make decisions based on an unobserved ability when all they can see is the race, gender and age of their counterparts. Assuming individuals are Bayesian learners, the proposed framework models agents as having prior beliefs or stereotypes about the ability levels for each observed characteristic (or combinations of characteristics.) Individuals also observe a noisy signal about each player's productivity. After observing the signals agents have to decide which person has the lowest (or highest) ability.

A difference in prior beliefs for each observed characteristic is the model's measure of discrimination. People who ignore the signals and keep their original stereotypes unaltered are considered as discriminating based on preferences. These people refuse to change their beliefs so their discriminatory behavior is not based on the revealed information. On the contrary, those who change their beliefs are considered as discriminating based on information.

I test this theory using a quasi-experimental setting. The data comes from the South African version of the television show *The Weakest Link*. In this show nine strangers compete for a winner-take-all prize. The prize increases with the number of correct answers. Players vote off one contestant at the end of each round. This paper shows how the data from the *The Weakest Link* can be used to estimate player's willingness to change their initial beliefs, therefore allowing us to identify adequate policies to reduce discrimination in post-apartheid South Africa.

Identifying the source of discrimination will help policy makers in two key areas. First, it will suggest better policies to reduce discrimination, complementing the efforts to eliminate poverty traps. Second, these policies will have a positive impact on the empowerment of non-whites, women, and other groups who suffer from discrimination.

The preliminary results show no discrimination by gender. With respect to race, players' negative stereotypes are against white opponents, however players are willing to change their beliefs after observing good signals. The opposite occurs regarding non-white opponents, that is, the initial stereotype remains the same regardless of these opponent's performance.

Players are not willing to change their beliefs about non-white opponents.

Following this introduction, the paper is divided in seven more sections. Section two briefly reviews previous measures of discrimination. Section three presents the model and its testable implication to distinguish between theories of discrimination. The data is described in section four and the estimation strategy is presented in section five. The main findings of the paper are shown in section six, followed by future extensions contained in section seven. Section eight summarizes the findings of this version of the paper.

2 Previous measures of discrimination

Measuring discrimination is a difficult task. During the apartheid era in South Africa and before the Civil Rights movement in the United States, there were laws that separated groups of the population. The discourse in the employment ads during those times also shows clear evidence of discrimination (Darity and Mason 1998). The current absence of these events is an improvement but does not imply discrimination has ended, rather that it is more subtle.

In economics, a common approach to measure discrimination is to decompose differences in wages (or in labor force participation) for two groups into observed factors and unobserved factors using the Oaxaca-Blinder decomposition. The observed factors include schooling and experience in the labor force as well as the returns of these variables. The unobserved factors are used as a proxy for discrimination. This methodology has been used in both developing and developed countries[4] and has the advantage of using household-level datasets, allowing researchers to draw conclusions about the population. However, this approach has been criticized as an inadequate approximation for discrimination because discrimination can also affect the observed factors such as schooling and experience in the labor force (Altonji and Blank 1999). Thus, the unobserved differences might not capture the full extent of discrimination.

Several alternatives have been explored to avoid this problem by using data from less conventional sources. The goal of this literature is to measure discrimination directly. For example, Ayres and Siegelman (1995) created audits where trained individuals from different races and genders bargained for a new car. The authors' findings suggest that dealers quoted

[4]See for example Altonji and Blank (1999) for applications in the U.S. and Lam and Leibbrandt (2004) and Casale (2003) for examples about South Africa.

lower prices for whites than blacks or female buyers using identical scripted bargaining strategies. Goldin and Rouse (2000) evaluate the impact of "blind" auditions on female musicians in orchestras. They found that females have a much higher probability to move to higher rounds of the auditions when performing behind a screen. The paper by Bertrand and Mullainathan (2004) mentioned in the introduction also falls in this category.[5]

Three conclusions can be drawn from the literature searching for evidence of discrimination. First, in order to test for discrimination, scholars are moving away from traditional household-level datasets. The studies mentioned above are closer to case studies and hence cannot make inferences about the entire population. The advantage though, is to have a clearer way of finding evidence of discrimination. Experiments are a good alternative to household-level datasets in this field. Second, unlike the Oaxaca-Blinder approach, the study of discrimination using these new methods has focused mostly on the United States, with almost no evidence from developing countries[6]. Third, all of these studies, including those using the Oaxaca-Blinder approach from developed and developing countries, are mute with respect to cause of discrimination. When evidence of discrimination is found, we do not know the reasons driving this behavior.

Economists have two main theories to explain why people discriminate. The first theory comes from Becker (1971). Becker explains discrimination as related to individual's preferences or taste. These individuals prefer not to interact with those discriminated against; hence, this approach has been labeled "preference or taste-based discrimination." Providing these individuals with information about the productivity of those suffering from discrimination will not change their discriminatory behavior. They behave as if they are unwilling to change their prior beliefs or negative stereotypes. The second theory comes from the simultaneous but independent work of Arrow (1973) and Phelps (1972). This approach "can be thought of as reflecting not tastes but perceptions or reality." (Arrow 1972, p. 23.) Here people use group identity, such as race, gender or age, as a proxy for unobserved ability. Because this approach relies on the information available to employers, it has been labeled "information-based" discrimination. In this case, agents do change their prior beliefs if we provide them with enough information about the productivity of those discriminated against.

[5]See Anderson, Fryer, and Holt (2005) for a survey of experiments measuring discrimination.

[6]Moreno, Ñopo, Saavedra, and Torero (2004) provides preliminary evidence of audit studies in Peru. See also Frijters (1999) for South Africa.

Distinguishing between the two theories is relevant in order to design policies to reduce discrimination. For example, to decide whether an affirmative action policy should be temporary or permanent it is crucial to know whether people's negative stereotypes about discriminated agents are subject to change. If agents are willing to learn, as in the case of discrimination based on preferences, a temporary affirmative action policy is needed to provide gains to those suffering from discrimination. Otherwise, when agents are unwilling to learn, these policies need to be permanent (Coate and Loury 1993). The model to test for the willingness to change stereotypes is presented next.

3 The model and testable implications

3.1 A model of learning

The model presented here is a one of Bayesian learning that relies on Arrow (1973) and Coate and Loury (1993). The idea is to have a group of individuals who do not know each other but have to make decisions based on the other agents' unobservable characteristics (such as productivity or ability). This lack of information makes individuals approximate these unobservable characteristics based on observable characteristics (such as race, age and gender.) Agents also observe a "noisy signal" that is imperfectly related to productivity.

The Bayesian part of the model comes from the assumption about how agents learn. First, for each observable characteristic, say, race, agents have a prior belief or stereotype about the proportion of non-whites with "high" or "low" productivity and a corresponding belief for whites. Agents have a probability distribution for the likelihood to observe a "good" signal from a non-white with low (or a high) productivity, and similar distribution for whites. All of these beliefs and probability distributions are predetermined and embedded in the participant's minds before meeting the other agents. Then information is revealed. Each agent observes a noisy signal from each participant. Using this signal, together with the priors and the probability distributions, each player constructs the posterior belief that a person has a low or high ability following Bayes theorem.

Formally, let I be the number of individuals, indexed by i and let k be the index of the other agents except i, which I will call "i's opponents" so if $i = 1$, then $k = \{2, 3, \ldots, I\}$ Let $j = \{1, \ldots, J\}$ be the set of observable characteristics. For example, when j refers to

gender then $j = 1$ refers to males and $j = 2$ to female. Let θ_j represent the unobservable characteristic of a person in group j and assume that θ_j is binary: $\theta_j = 1$ when ability is high and $\theta_j = 0$ when ability is low. I am also assuming that there is no heterogeneity within each value of θ_j. Also, y_{jk} denotes the observed quality of the signal from opponent k that belongs to group j. When the signal is "good" $y_{jk} = 1$, otherwise $y_{jk} = 0$.

We turn now to the prior beliefs. Let $\text{Prob}_i(\theta_j = 0) = \alpha_{ij}^0$ be the probability that person i's belief is that players from group j have low ability. To save on notation let me erase the i-subscript, so I will refer to α_j^0 instead. Let p_j be the probability that player i relates a "good" signal ($y_{jk} = 1$) as coming from a high ability player from group j, and let q_j be analog for a low ability player. Assume that the noisy signal is observed S times, so the number of good signals is given by $Y_{jk} = \sum^S y_{jk}$. The probability of observing Y_{jk} follows the below binomial distribution when a person has high ability:

$$\text{Prob}_i(Y_{jk}|\theta_j = 1) = \frac{S!}{Y_{jk}!(S - Y_{jk})!} p_j^{Y_{jk}} (1 - p_j)^{S - Y_{jk}} = h_{ip}(Y_{jk}) \tag{1}$$

and similarly for a person with low ability:

$$\text{Prob}_i(Y_{jk}|\theta_j = 0) = \frac{S!}{Y_{jk}!(S - Y_{jk})!} q_j^{Y_{jk}} (1 - q_j)^{S - Y_{jk}} = h_{iq}(Y_{jk}) \tag{2}$$

The assumption of Bayesian updating defines the way agents update their beliefs. Let α_{jk}^1 be the probability that player i assigns to her k-th opponent, belonging to group j, as having a low ability after observing k's signals summarized by Y_{jk}. Formally,

$$\begin{aligned}
&\text{Prob}_i(\text{player k} \in j \text{ is low type } |k \text{ has } Y_{jk} \text{ correct questions}) \\
&\quad = \text{Prob}_i(\theta_j = 0|Y_{jk}) \\
&\quad = \frac{\text{Prob}_i(Y_{jk}|\theta = 0)\text{Prob}_i(\theta_j = 0)}{\text{Prob}_i(Y_{jk}|\theta_j = 0)\text{Prob}_i(\theta_j = 0) + \text{Prob}_i(Y_{jk}|\theta_j = 1)\text{Prob}_i(\theta_j = 1)} \\
&\quad = \frac{h_{iq}(Y_{jk})\alpha_j^0}{h_{iq}(Y_{jk})\alpha_j^0 + h_{ip}(Y_{jk})(1 - \alpha_j^0)} \\
&\quad = \alpha_{jk}^1(\alpha_j^0, p_j, q_j; Y_{jk})
\end{aligned} \tag{3}$$

where $h_{ip}(\cdot)$ and $h_{iq}(\cdot)$ are functions defined in equations (1) and (2) respectively. The pos-

terior probability α^1_{jk} is then a function of the structural parameters of the model (α^0_j, p_j, q_j) for all j, as well as the information revealed from each opponent in the form of Y_{jk} for all k and all j.

Note that when $p_j = q_j$ for some j equations (1) and (2) are identical. Hence in equation (3) the functions h_{iq} and h_{ip} cancel out from the numerator and denominator. What is left in equation (3) is α^0_j in the numerator and $\alpha^0_j + 1 - \alpha^0_j$ in the denominator. This in turn implies that the posterior probability α^1_{jk} is equal to the prior probability α^0_j for all $k \in j$.

3.2 Testable implications

A test for discrimination can be developed by observing differences in prior beliefs. If person i believes that an opponent from group j has a higher probability to be a low ability player than an opponent from group t, one might say that is evidence of discrimination against members of group j. This is captured in the following definition

Test 1 (Test for discrimination) *If $\alpha^0_j > \alpha^0_t$ for some $t \neq j$ then members of group j are discriminated against.*

To distinguish between theories of discrimination we rely again on the values of p_j and q_j and consider the extreme case described above, where $p_j = q_j$. We now can present our way to distinguish between theories of discrimination:

Test 2 (Unwillingness to change) *Assuming we found evidence of discrimination against members of group j, in Test 1, then if $p_j = q_j$ the discrimination is based on preferences.*

As showed above, if for some j we have $p_j = q_j$ it implies that the prior and the posterior are the same. In this case, agents behave as if they refuse to update their beliefs. The revelation of information, through the noisy signal, does not affect their decision. The discrimination is not related to information, it is a matter of preferences. This is reasoning behind Test 2. It is then straight forward to show how to test for discrimination based on information, as the alternative hypothesis to Test 2.

Test 3 (Willingness to change) *Assuming we found evidence of discrimination against members of group j in Test 1, then if $p_j \neq q_j$ the discrimination is based on information.*

Under Test 3 agents do update their beliefs, hence if we provide them with enough information about their opponents' performance their priors will change. Agents for whom $p_j \neq q_j$ discriminate because they do not have enough information about their opponents, not because they refuse to learn [7].

This simple model of Bayesian learning creates a framework to test for discrimination but also to distinguish between theories of discrimination. The data to be used to test this theory in post-apartheid South Africa is presented below.

4 Data sources

The model presented above is simple but general enough so it can be applied in different scenarios where agents receive a noisy signal and have chance to update their priors. In addition to this we need a setting where agents have a chance to reveal their posterior beliefs. This is a hard task because it requires a particular set up. Hence to test the theory I use data from the South African version of the TV show *The Weakest Link*.

The Weakest Link is a winner-takes-all television game where nine contestants answer several trivia questions. These contestants have a decreasing amount of time to answer as many questions as possible in each round. At the end of a round, each player decides individually, secretly and simultaneously who to vote off the game. When the votes are revealed, the person with the highest number of votes against them leaves the game. The remaining contestants keep answering questions and eliminating one player per round until two players are left. Within these two players, the one who in the final stage answers the most questions correctly wins. The prize is a function of the number of correct questions throughout the game[8] They can win a maximum of R60,000, approximately US$10,000 or US$ 21,200 using PPP.

The game has all the components needed to estimate the model in section 3.1. First, the participants do not know each other before playing the game, which is a requirement for players to have priors based on observables believable. Second, players are drawn from all groups in the South African population in terms or race, gender and age (18 and above).

[7] I section six I explore, for those who update their priors, how fast or slow they update in order to have a broader sense of the distinction between the two competing theories.

[8] The prize is the amount of "banked" money, and baking is allowed after a correct answer.

Third, players have to identify their opponents' ability to find out who is the weakest link (i.e. the player with the lowest ability) and who is the strongest link (i.e. the player with the highest ability). Fourth, this "ability" is observed as a noisy signal in the form of the number of questions answered correctly. Answering a question is considered a noisy signal because the questions' difficulty is not homogenous. The observed performance of each player becomes a random variable. Fifth, at the end of each round, players reveal their posterior probabilities by their voting patterns which in principle we can assume reflects their choice regarding who they think is the weakest link.

Another advantage of using this game is the prizes which are much higher than the ones used in experiments. The costs of using this game are similar to the ones found in the new literature detecting discrimination as described in section one. The sample is not nationally representative and the demographics do not necessarily match the population distribution. However, I believe the benefits exceed the costs. The possibility of applying methods that were -up to now- applied to developed countries, along with the availability of a methodology that allows for the identification of the underlying reason for discrimination is worth pursuing. The benefits increase as this paper analyzes the case of post-apartheid South Africa.

Note that the approach used here differs from the methodology used by Levitt (2003) and Antonovics, Arcidiacono, and Walsh (2004). These authors used data from the U.S. version of the show to distinguish between theories of discrimination by analyzing the dynamic feature of the game. They argue that players would find it optimal to eliminate the weakest players in the early rounds because the prize increases with the number of correct answers. Towards the end of the game the desire to win might become more important for players and they will vote off the strongest players instead. However, the dynamics might imply strategies contrary to the ones argued in their papers due to issues such as reputation, punishments and disclosure of information. Here I focus on the feature of the game related to the development of information whose consequences are formally presented in the model of section 3.1. To avoid the role of history, I use the voting patterns from the first round only[9].

[9]For example, vengeance can be a motive for a player to vote off an opponent who voted against her in previous round. Also, from round two onwards the player with the highest number of correct questions in the previous round start the next round, so it is made public who the strongest player is after each round. Finally, once the votes are made public (and before asking the voted off person to leave the game) the show's host interviews two or three participants (at her discretion) asking them about their reasons for their vote

The data is collected from videotapes of three seasons. I prepared a questionnaire to capture the data (available upon request.) There are 16-18 shows per season, once we exclude the shows where celebrities play for charity. With three seasons so far, the sample size has around 450 players. Season three is currently being broadcasted in South Africa. This season will run until the end of May. For this version of the paper I will use data only from the first season. The main variables and their basic statistics are shown in Table 1 below.

Table 1: Basic Statistics

Variables	Type	Nobs.	Mean	Std. Dev.	Median	Min.	Max.
White	Binary	135	.733	.444	1.00	.000	1.00
Male	Binary	135	.533	.501	1.00	.000	1.00
Age	Years	135	34.95	11.01	31.00	19.00	74.00
Johannesburg	Binary	135	.400	.492	.000	.000	1.00
Traditional cloth	Binary	135	.015	.121	.000	.000	1.00
Questions	Number	135	2.70	.519	3.00	2.00	4.00
Correct answer	Number	135	1.91	.851	2.00	0.00	4.00
Correct answers	Proportion	135	.710	.288	.667	.000	1.00
Voted off	Binary	135	.370	.485	0.00	0.00	1.00
Votes against	Number	135	1.00	1.82	0.00	0.00	8.00

There are 135 participants in the first season of the show. Whites are over represented in the sample and they account for almost three quarters of the players. However, the sample is almost evenly distributed in terms of gender. The players's age ranges from 19 to 74 years old. Most players are from the Johannesburg-Pretoria area which is the financial center of South Africa. Very few of them wore "traditional clothes." Players have two minutes and 50 seconds to answer as many questions as they can in the first round, so the total number of questions varies by participants[10]. The median player answers three questions and very few answer two or four questions. On average players answer two questions correctly and

changing which might change the information set of the remainder participants. None of this occurs in round one.

[10]Hence S is not fixed and for the estimation we use S_i in equations (1) and (2).

the proportion of correct answers is just above 70%. The last two rows in Table 1 show that more than a third of the participants were voted off, that is, they received at least one vote against, however the distribution of the number of votes received is skewed to the left. The next section explains how to use this data to evaluate the model presented in section three.

5 Estimation

To estimate the structural parameters of the model $(\alpha_j^0, p_j, q_j;$ for all $j)$ I will use maximum likelihood methods. Let $d_{ijk} = 1$ if individual i votes against opponent k that belong to group j and $d_{ijk} = 0$ otherwise. Under the assumption that players find it optimal to eliminate the weakest link in the first round the likelihood function is given as follows

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha^0}, \boldsymbol{p}, \boldsymbol{q}; \boldsymbol{Y}) &= \prod_i^N \prod_{k \neq i} \Big(\text{Prob}_i(\text{vote off player } k \in j)^{d_{ijk}} \\ &\quad \times (1 - \text{Prob}_i(\text{vote off player } k \in j))^{1-d_{ijk}} \Big) \\ &= \prod_i^N \prod_{k \neq i} \alpha_{jk}^1(\alpha_j^0, p_j, q_j; Y_{jk})^{d_{ijk}} (1 - \alpha_{jk}^1(\alpha_j^0, p_j, q_j; Y_{jk}))^{1-d_{ijk}}\end{aligned} \tag{4}$$

where α_{jk}^1 is the function defined in equation (3) and N is total number of players, also parameters in **bold** reflect vectors.

To test for discrimination and to distinguish between theories of discrimination I will use the Likelihood Ratio, where the hypotheses are described in Tests 1 to 3 above. The Likelihood function described in equation (4) is highly non-linear due to the binomial distribution of Y_{jk} together with the Bayes rule to update the posterior probability in equation (3). In Proposition 1 below, I show that all the parameters are identified because α_{ijk}^1 is not a linear function of the structural parameters.

Proposition 1 (Identification) *The set of parameters* $\boldsymbol{\Psi} = (\boldsymbol{\alpha^0}, \boldsymbol{p}, \boldsymbol{q})$ *is fully identified.*

Proof. Suppose not. If the parameters are not identified, adding a non-zero vector $\boldsymbol{\delta}$ in equation (3) would not affect its value. This is equivalent to have $\alpha_{ijk}^1(\boldsymbol{\Psi} + \boldsymbol{\delta}) = \alpha_{ik}^1(\boldsymbol{\Psi}) +$

$\alpha^1_{ijk}(\boldsymbol{\delta})$ which in turns requires $\alpha^1_{ijk}(\cdot)$ to be a linear function. But as shown by equation (3), $\alpha^1_{ijk}(\cdot)$ is not linear with respect to α^0_j, p, q which contradicts the initial statement. ■

The structural parameters α_j, p_j, q_j, for all j, are probabilities and hence limited to take values between zero and one. To insure that I assume that each can be expressed as a logistic transformation. Let $m = \{\alpha_j, p_j, q_j\}$, then the parameters will be estimated according to this equation:

$$m = \frac{e^{\gamma_m}}{1 + e^{\gamma_m}} \tag{5}$$

6 Results

6.1 Performance and voting patterns

As mentioned in section four, player performance is not homogenous. Figures (1) and (2) show how the proportion of correct answers varies by race and gender, respectively. White players have a higher probability of getting more questions right and there is no major difference in the distribution when comparing genders. Table 2 below complements the findings of these figures.

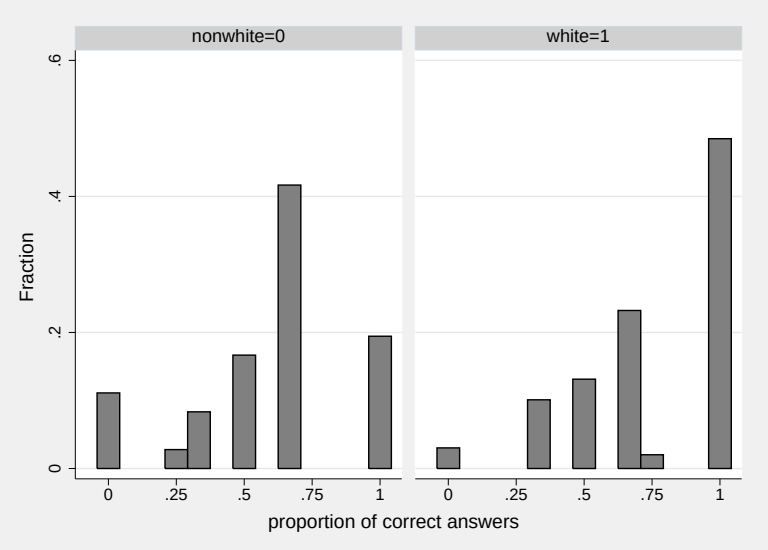

Figure 1: Performance by race

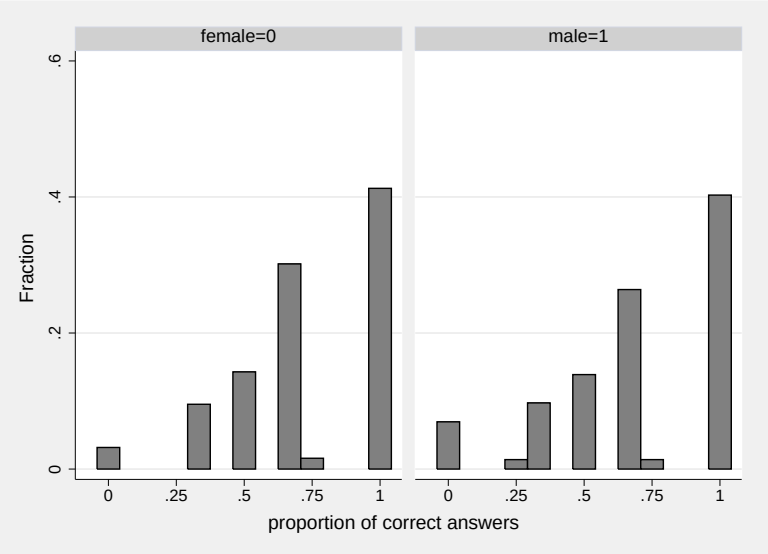

Figure 2: Performance by gender

Column (1) relates players characteristics with their performance measured by whether the player was the weakest link (i.e. the player with the worst performance) or not. None of the players' characteristics correlates with such a measure of performance. In column (2) we measure performance as the proportion of correct answers. We find that white and older players have a lower probability of being the weakest links. Gender has no role explaining this measure of performance.

Columns (3) to (8) refer to the voting patterns. Column (3) shows that the probability of receiving at least one vote is higher for non-white players. Since they also have a lower performance in columns (4) and (5) we control for that by including the two previous definitions of performance. Once we do that there are no differences in being voted off by the race of the player, but age shows a significant correlation with being voted off. Similarly, column (6) shows that the number of votes received is higher for non-whites, but the racial difference disappears after controlling for performance. This might suggest that players are not indifferent to the information revealed in the game.

Table 2: Performance, votes and observable characteristics

Model :	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	-0.099	n.a	-0.378	-1.039	1.218	0.852	-0.466	1.517
	(0.550)		(0.409)	(0.457)	(0.546)	(0.320)	(0.360)	(0.308)
Male=1	0.265	-0.006	0.054	-0.046	0.070	0.121	-0.122	-0.115
	(0.278)	(0.027)	(0.227)	(0.246)	(0.276)	(0.177)	(0.187)	(0.181)
White=1	-0.352	0.074	-0.523	-0.426	-0.280	-0.434	-0.182	-0.071
	(0.284)	(0.032)	(0.260)	(0.282)	(0.305)	(0.188)	(0.198)	(0.191)
Live in	-0.161	-0.002	-0.043	-0.011	-0.080	-0.291	-0.190	-0.019
Johannesburg	(0.280)	(0.027)	(0.231)	(0.248)	(0.279)	(0.186)	(0.187)	(0.192)
Age (yrs)	-0.024	0.003	0.012	0.022	0.039	-0.015	0.001	0.018
	(0.017)	(0.001)	(0.010)	(0.011)	(0.013)	(0.009)	(0.009)	(0.009)
Weak player				1.950			2.036	
				(0.417)			(0.183)	
Prop. of correct					-4.036			-3.719
answers					(0.651)			(0.320)
No. obs :	135	1351	135	135	135	135	135	135
(Pseudo) R^2	0.066	0.067	0.028	0.192	0.362	0.034	0.286	0.366

Column (1) Probit: Weakest player, marginal effects
Column (2) Tobit: Proportion of correct answers, marginal effects
Columns (3)-(5) Probit: Voted off, marginal effects
Columns (6)-(8) Poisson: Number of votes received
Note: Standard deviations in parenthesis.

6.2 Discrimination and its sources

We now present the estimates for the structural parameters of the model to test for discrimination and it sources. In this version of the paper I will focus on the racial and gender differences in votes, leaving the analysis of discrimination by age for future versions of the paper. Table 3 and 4 show the estimates for the parameters α_j^0, p_j, q_j when $j = \{$men,women$\}$ and when $j = \{$whites, non-whites$\}$, respectively.

Table 3: Estimates for gender discrimination, by groups

	Sample of voters				
Parameters	All	Whites	Non-whites	Men	Women
α^0_{women}	.028	.019	.300	.042	.030
	(.027)	(.007)	(.346)	(.052)	(.013)
α^0_{men}	.086	.074	.139	.151	.048
	(.058)	(.059)	(.177)	(.132)	(.050)
p_{women}	.970	1.00	.630	.918	1.00
	(.091)	(1.5E-4)	(.323)	(.113)	(3E-4)
p_{men}	.823	.841	.736	.735	.892
	(.097)	(.101)	(.279)	(.167)	(.115)
q_{women}	.857	1.00	.199	.613	1.00
	(.391)	(8.1E-4)	(.230)	(.374)	(.001)
q_{men}	.419	.428	.358	.301	.560
	(.172)	(.197)	(.340)	(.192)	(.310)
Nobs.	135	99	36	72	63

Note: Standard deviations in parenthesis. Maximum likelihood estimates.

For both tables, the first column presents the estimates using the full sample. In the case of gender, the players' beliefs that men and women are low-ability players (i.e. α_j^0) is less than 10%, but are higher when looking at differences in races in table 4. In both tables, the point estimates for p_j (i.e. the probability that a good signal comes from a *high-ability* player) are bigger than q_j (i.e. the probability that a good signal comes from a *low-ability* player).

Also, there are interesting differences on how these perceptions change by groups in both tables, but they could be related to the precision of the estimates due to differences in sample size.

Table 4: Estimates for racial discrimination, by groups

	Sample of voters				
Parameters	All	Whites	Non-whites	Men	Women
α^0_{nw}	.021	.019	.156	.026	.017
	(.008)	(.009)	(.282)	(.031)	(.010)
α^0_w	.164	.147	.213	.163	.167
	(.100)	(.110)	(.217)	(.137)	(.146)
p_{nw}	1.00	1.00	.778	.995	1.00
	(1.2E-4)	(2.0E-4)	(.311)	(.132)	(1.8E-5)
p_w	.725	.753	.643	.760	.675
	(.119)	(.128)	(.283)	(.136)	(.219)
q_{nw}	1.00	1.00	.350	.979	1.00
	(6.1E-4)	(.001)	(.426)	(.566)	(1.1E-4)
q_w	.277	.279	.256	.263	.284
	(.124)	(.146)	(.240)	(.152)	(.209)
Nobs.	135	99	36	72	63

Note: Standard deviations in parenthesis. Maximum likelihood estimates. nw=non-whites, w=whites.

The estimated parameters of these two tables allow us to perform the tests described in section 3.1. We first test for the existence of discrimination following Test 1. The results are presented in table 5. We cannot reject the null hypothesis that the prior beliefs (α^0_j) are the same for both genders, suggesting no evidence of gender discrimination. This results holds for all sub-samples.

The findings are different when we look for racial discrimination. The bottom panel of table 5 shows that we can reject the null hypothesis of equal priors for different races. This is evidence of discrimination by race. There two important issues to mention here. First, the existence of racial discrimination does not hold for all sample divisions (each column in table 5). Only whites and women voters exhibit evidence of racial discrimination, while

non-whites and men do not. However, as mentioned before we have to take into account the differences in sample size. Including the other two seasons of the show will help to identify this issue more accurately.

Table 5: Tests for discrimination

	Sample of voters				
Groups	All	Whites	Non-whites	Men	Women
Gender					
Statistic	.947	2.29	.195	.710	.181
p-value	.331	.130	.659	.399	.670
Race					
Statistic	7.31	4.83	.023	1.50	4.15
p-value	.007	.028	.879	.220	.042

$H_0 : \alpha_j^0 = \alpha_t^0$. Critical value: $\chi^2_{95\%}(1) = 3.84$

The second important issue is related to the identification of the racial group that is discriminated against. Table 4 shows that the prior probability that whites are low-ability players is higher, not lower, than the corresponding prior for non-whites. This suggest that those who are discriminated against are the white players, not the non-white players. This feature is found also when looking at the estimates from white and women voters. In section seven I discuss the appropriateness of this test for discrimination[11].

In table 6 we present the results of the tests to identify the sources of discrimination. Since we only find racial discrimination we limit the tests to whether players are willing to

[11]It is possible that this result might be caused by a sample selection problem. To appear in the show people need to send an application and the producers of the select a smaller sample out of the pool of applicants. It might be the case that the discrimination against white reflect the characteristics of the sample. I plan to interview the producers of the show to learn more about the selection of the contestants, to understand better this issue. A second explanation for this surprising result could be associated with the strategy followed by the players. I assumed that players find it optimal to eliminate the weakest player because the having better players increases the chances of having a higher prize at the end of the game. But players might find optimal to do the opposite: vote off the strongest players because they also want to win the game. In section seven I present a test to evaluate which strategy is represents the data better. Also, in the appendix I show the conditions for players to select each strategy.

change their negative stereotypes regarding white and non-white opponents. The results suggest that players behave as if they are willing to change their initial beliefs with respect to whites but not with respect to non-whites. We strongly reject the null hypothesis that $p_j = q_j$ for the former but not for the latter. This result holds across all subgroups where we found evidence of discrimination, but also for men. Non-white voters, however, do not discriminate as stated before but also are willing to update their beliefs about both types of players: whites and non-whites.

Table 6: Testing theories of discrimination: race

	Sample of voters				
Groups	All	Whites	Non-whites	Men	Women
Non-whites					
Statistic	2.7E-4	1.2E-5	6.88	.001	3.8E-4
p-value	.987	.997	.009	.970	.985
Whites					
Statistic	127	87.7	22.7	77.6	44.0
p-value	.000	.000	.000	.000	.000
$H_0 : p_j = q_j$. Critical value: $\chi^2_{95\%}(1) = 3.84$					

These results imply that while players' initial beliefs favor non-whites, they are not willing to change that belief in the presence of good or bad information. Consider, for example, the case of an average player facing two opponents: one white and one non-white. This player believes that the probability that his white opponent is a low-ability contestant is 0.164 (from table 4) and 0.021 for his non-white contestant. If both players answer all three questions correctly and given the estimates for p_j and q_j for each race, the posterior for the non-white opponent will remain the same (0.021) but for the white opponent, the posterior falls to 0.011 which is almost half of the one for non-whites. Non-white players with good signals will be eliminated instead of white players with the same good signals. The next section describes four extensions to be implemented in future versions of this paper.

7 Future extensions

The extensions contained in this section will be part of future versions of the paper.

7.1 Observed heterogeneity

So far to account for heterogeneity the sample was divided into different groups as shown in the tables above. The main drawback of this approach is the lack of testing for differences in behavior by the characteristic of the voters. The model can be extended to include complementary demographic characteristics of the decision makers but also to allow for characteristics of their opponents. Let $\boldsymbol{x_i}$ and $\boldsymbol{z_{jk}}$ be the vectors of demographics characteristics of player i and her k-th opponent, respectively. Equation (5) can be modified so $\boldsymbol{\gamma_m} = \boldsymbol{x_i'\beta_m} + \boldsymbol{z_{jk}'\lambda_m}$ making $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ shifters affecting the structural parameters.

7.2 Unobserved heterogeneity

It is possible also to allow for unobserved heterogeneity by assuming that for each player the prior is a random draw from a known distribution $F(\alpha_j^0)$. In this case the posterior can be expressed as follows:

$$\alpha_{ijk}^1 = \int \frac{h_{iq}(Y_{jk})\alpha_j^0}{h_{iq}(Y_{jk})\alpha_j^0 + h_{ip}(Y_{jk})(1-\alpha_j^0)} dF(\alpha_j^0) \tag{6}$$

this expression for the posterior replaces the one in equation (3). Once this new expression is included in the likelihood function (Eq. 4) we can use Bayesian methods to estimate the posterior.

7.3 An alternative test of discrimination

Let us return now to the way to test for discrimination. In Test 1 differences in prior beliefs is taken as evidence of discrimination. However, it might be the case that a higher prior reflects an accurate knowledge of the environment rather than discrimination. Consider the case when j indexes races applied to South Africa. A well informed agent who knows that Africans received a lower quality education during the apartheid era might have a $\alpha_{j=african}^0$ higher than for other races, reflecting her beliefs that African participants have a higher

probability of having low ability due the restrictions faced during the apartheid regime. It is possible then to have different priors for each j without discriminating, necessarily, against any group[12]. This might explain why we found discrimination against white using our initial definition of discrimination. An alternative definition of discrimination can be found below in Test 4.

Test 4 (Alternative test for discrimination) *If either $q_j < q_t$ or $p_j < p_t$ for any $t \neq j$ then members of group j are discriminated against.*

The idea behind this definition of discrimination is that, conditional on the (unobserved) ability, agents of all groups are identical. If two opponents, one from group j and the other from group t have a low ability, the probability of having a good or a bad signal (y_{jk}) from each of them must be the same in the absence of discrimination because there is no heterogeneity within each level of ability. Note that this result does not arise from the discrete values of the ability parameter θ. This result holds even in the case when θ is a continuous variable.

7.4 Players' strategy

Finally, a key assumption throughout the paper was that during the first round of the game the players' strategy is to vote off the weakest player. It is possible to test for the accuracy of this assumption. To do that the likelihood function in equation (4) can be modified to allows for different rules (voting the weakest or the strongest link) in the voting pattern and hence identify in the data which strategy described the observed behavior better.

$$\begin{aligned}\mathcal{P}_{ikj} &= \text{Prob}_i(\text{vote weak player}|\text{rule 1})\text{Prob}_i(\text{rule 1}) \\ &+ \text{Prob}_i(\text{vote strong player}|\text{rule 2})\text{Prob}_i(\text{rule 2}) \\ &= \alpha^1_{ijk}(\cdot)\frac{e^\tau}{1+e^\tau} + (1-\alpha^1_{ijk}(\cdot))\frac{1}{1+e^\tau}\end{aligned} \tag{7}$$

Using equation (7) the likelihood function changes to:

[12]However, notice that for other characteristics this might not be the case.

$$\mathcal{L}(\boldsymbol{\alpha^0}, \boldsymbol{p}, \boldsymbol{q}; \boldsymbol{Y_{jk}}) = \prod_{i}^{N} \prod_{k \neq i} \mathcal{P}_{ikj} \tag{8}$$

Future versions of the paper will include this test.

8 Conclusions

"Not to know is bad. Not to wish to know is worse." -Wolof proverb.

This paper proposes a new test to distinguish between theories of discrimination: preferences versus information. Using a model of learning, players are assumed to have negative stereotypes or prior beliefs about their opponents' unobserved productivity. Discrimination based on preferences takes place when players behave as if they refuse to update their initial beliefs. Those who do change their beliefs are seing as discriminating based on information.

The preliminary results from South Africa described in this paper show no evidence of discrimination against women, but the evidence regarding racial discrimination is more complex. Players have a prior belief that benefits non-whites instead of whites but they are reluctant to change their impression for the former. Good or bad signals about non-white opponents have no impact on players' behavior. The contrary is found for the case of white opponents. A good white opponent will have a lower probability of being voted off compared to a good non-white opponent.

Several extensions have been mentioned in section seven and they will be part of future versions of this paper. The current results, however, have severe implications in terms of the permanency of affirmative action policies. The fact that players are not willing to update their stereotypes about non-whites could be used as an argument for long-term affirmative action policies.

References

Altonji, J., and R. Blank (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, pp. 3144–3259, Amsterdam. North-Holland.

Anderson, L. R., R. G. Fryer, and C. A. Holt (2005): "Discrimination: Experimental Evidence from Psychology and Economics," in *Forthcoming Handbook on Economics of Discrimination*, ed. by W. Rogers.

Antonovics, K., P. Arcidiacono, and R. Walsh (2004): "Games and Discrimination: Lessons From The Weakest Link," Working paper, University of California at San Diego.

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

Ayres, I., and P. Siegelman (1995): "Race and Gender Discrimination in Bargaining for a New Car," *American Economic Review*, 85(3), 304–21.

Becker, G. S. (1971): *The Economics of Discrimination*. University of Chicago Press, Chicago, 2nd edn.

Bertrand, M., and S. Mullainathan (2004): "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 94(4), 991–1013.

Carter, M., and C. Barret (2005): "The Economics of Poverty Traps and Persistent Poverty: An Asset-Based Approach.," Working paper, Cornell University.

Casale, D. (2003): "The Rise in Female Labour Force Participation in South Africa: An Analysis of Household Survey Data, 1995-2001," Department of economics, University of Natal.

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

Darity, William A, J., and P. L. Mason (1998): "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender," *Journal of Economic Perspectives*, 12(2), 63–90.

Frijters, P. (1999): "Hiring on the Basis of Expected Productivity in a South African Clothing Firm," *Oxford Economic Papers*, 51(2), 345–54.

Goldin, C., and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 90(4), 715–741.

Lam, D., and M. Leibbrandt (2004): "What's happened to inequality in South Africa since the end of apartheid?," Manuscript, University of Cape Town.

Levitt, S. D. (2003): "Testing Theories of Discrimination: Evidence from "Weakest Link"," NBER Working Papers 9449, National Bureau of Economic Research, Inc.

Moreno, M., H. Ñopo, J. Saavedra, and M. Torero (2004): "Gender and Racial Discrimination in Hiring: A Pseudo Audit Study for Three Selected Occupations in Metropolitan Lima," Discussion Paper 979, Institute for the Study of Labor (IZA).

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62(4), 659–61.

A Strategies

Consider the simple game where there are three players and they have to vote-off one player. The remainder two players compete for the prize where the winner takes it all. Player i has to decide who to vote off: player A or player B. If A is voted-off, two situations can occur. With probability $1-\omega_b$, player i losses and gets zero. With probability ω_b she wins and her expected payoff is $\delta_b z_1 + (1-\delta_b)z_2$ where z_1 and z_2 are the low and high payoffs respectively, and δ_b reflects the probability of achieving the low payoff z_1.

If B is voted off, i will face player A in the next round where the payoffs are the analogs to the above ones with probabilities ω_a and δ_a instead. Player i's choice are depicted in Figure 3.

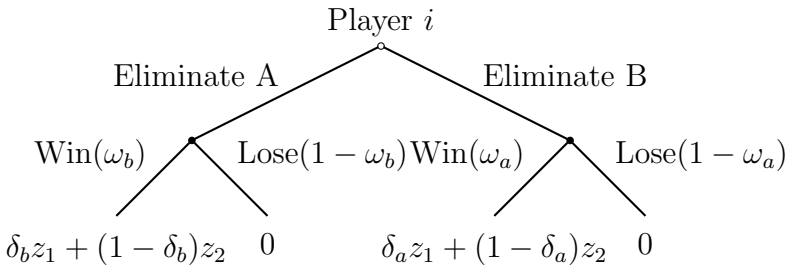

Figure 3: Player i's choices

Let us now assume that $\omega_b > \omega_a$ and that $\delta_b > \delta_a$. The former states that player i has a bigger chance of winning if facing player B and the latter implies that the high prize is more probable when playing with player A. These assumptions imply that player A is a better player than B. It is more difficult to win when playing against her but the expected prize increases.

If player i decides who to vote off by maximizing her expected value she will vote off player A if and if:

$$\frac{\omega_b}{\omega_a} \geq \frac{\delta_a(1-\frac{z_1}{z_2})-1}{\delta_b(1-\frac{z_1}{z_2})-1} \tag{9}$$