Lies, Discrimination, and Internalized Racism: Findings from the lab.

David, Wozniak and Tim, MacNeill

Eastern Michigan University, University of Ontario Institute of Technology

25 October 2015

Lies, Discrimination, and Internalized Racism: Findings from the lab.[1]

David Wozniak[2a] and Tim MacNeill[b]

[a]Eastern Michigan University: Department of Accounting and Finance

[b]University of Ontario Institute of Technology: Faculty of Social Sciences and Humanities

October 25, 2015

[1]Funding for this project came from Eastern Michigan University: New Faculty Grant. We also thank participants from the Regional ESA Meetings (Tucson, AZ) 2012 and seminar participants at UCSD. All errors are our own.

[2]Corresponding author: Tel. 1.734.545.8715 Fax. 1.734.482.0806 Email address: dwozniak@gmail.com

Abstract

We simulate a job application/hiring market in the lab to examine racial discrimination. We find little evidence of ability differences based on race but we find taste-based racism between groups and statistical racism within groups. When candidates are given the opportunity to lie about their abilities, all groups discriminate against Blacks, suggesting statistical discrimination. But Whites continue to discriminate against Blacks when actual abilities of the candidate are known, suggesting taste-based discrimination. In contrast to the bulk of studies that attempt to establish racism in general as either a taste-based or statistical, our design allows us to show that the type of discrimination can depend on the personal characteristics of the discriminating individual along with the contextual information available.

1 Introduction

Despite efforts taken to eliminate legal forms of discrimination, racial inequality remains an enduring problem in the 21st century. In the United States, for example, median yearly earnings of African American males are 25 percent lower than those of their White counterparts and 22.6 percent less of them are employed compared to Whites (BLS 2014). Recently, tensions have transformed into violence and protest in the American Midwest amidst claims of institutionalized racism, police mistreatment, and a related lack of opportunity for African Americans (Apel 2014). These current trends attest to the need for economic research on discrimination.

To help our understanding of discrimination, we create a simulated labor market with university students in a racially diverse area of the American Midwest. Subjects perform a task and choose/hire teammates based on a variety of characteristics such as pictures of individuals, stated task performance and real task performance. After these treatments, subjects choose partners based on a combination of stated performance and a photograph followed by the final treatment of real performance and a photograph. Using these treatments allows us to identify if racism exists whenever race is salient and whether it remains even when the actual ability of the candidate is known. The design is unique in that it allows us to identify racism within groups and across groups and allows us to avoid deception in fieldwork.

The experiment provides us with some important and consequential findings. Racism of Whites against Blacks is wholly taste-based and exists during all treatments where race is salient. We find that internalized discrimination is statistical and occurs only when individuals see stated performance versus real performance. Part of the observed discrimination occurs because both Blacks and Whites expect Blacks to lie more than Whites, while in fact they are more honest in our laboratory setting. This internalized racism by the discriminated group is removed with information about actual abilities. In contrast, discrimination of Whites against Blacks does not disappear when information on abilities is known. This suggests that discrimination is not simply a result of incomplete information for this group.

Studies on discrimination have been undertaken in economics for a significant period of time. The earliest studies questioned when, where, and if discrimination existed and discerned its impact on labor markets through the use of historical datasets (Fawcett 1892, Edgeworth 1922, Bergmann 1971). More recently, a move toward simulated labor markets produced robust findings which

matched those of earlier studies – that discrimination exerts a significant impact on hiring practices (Bertrand and Mullainathan 2004, Riach and Rich 2006, Rubinstein and Brenner 2013). Further experimental studies have found that discrimination is often implicit – unconscious– as opposed to being overt (Bertrand, Chugh and Mullainathan 2005). Although they have provided important and interesting results, such field experiments have been criticized for employing deception within real labor markets (Riach and Rich 2004). This ethical problem is avoided in the method of our study, since a laboratory-based labor market simulation is used in lieu of an actual labor market intervention.

Leaving the question of whether or not discrimination exists and if it matters relatively resolved, foundational studies by Becker (1971,2010) and Arrow (1973) moved away from the question of if discrimination exists, to why discrimination exists . Becker's argument was that discrimination was a taste-based phenomenon. In this view, people simply have a taste or distaste for other people depending on whether these others fall into particular racial, gender or other categories. Arrow, on the other hand, theorized discrimination as a statistical phenomenon - arguing that, in the absence of information about the abilities or trustworthiness of others, people tend to discriminate based on observable traits. This is a product of statistical assumptions that may be true of the affiliated group on average, but possibly not true of the individual in question.

Much of the recent work on discrimination has contributed to this debate. This work has helped the discussion to move away from either/or arguments about whether discrimination is taste-based or statistical, to a more nuanced focus on when and where discrimination may be animated from either of these sources. Castillo and Petrie (2010), for example, find evidence of statistical discrimination in a laboratory experiment regarding cooperation in the provision of public goods. In a later experiment, however, they find that statistical discrimination exists regarding race, but is taste-based regarding beauty (Castillo, Petrie and Torero 2012). Gneezy, List, and Price (2012) have expanded this investigation in a number of field experiments, finding that whether discrimination is statistical or taste-based is dependent of whether or not the discriminator believes the discriminated-against trait, such as race, gender, sexuality, or disability to be due to choice or not - with statistical discrimination prominent in the former instance, and taste-based in the latter. Rich (2014) conducted a meta-analysis of such studies finding that discrimination was statistical mainly in product markets and taste-based elsewhere .

We have good reason to expect to find statistical racism against Blacks in our study. There is a widely publicized performance gap between Blacks and Whites on mental and academic tasks in American society. This has been noted in terms of SAT scores (Card and Rothstein 2007, Fleming 2002), high school (Fletcher and Tienda 2015), and post-secondary performance (Men n.d.). In all of these cases, Blacks tend to perform worse than Whites. Assuming that participants are aware of these widely known performance gaps, we would expect to find a discrimination against the hiring of Black participants and for this to be driven by a lack of information on the part of the "employer" as to the actual competence of the field of potential candidates.

Researchers have recently begun to criticize studies on racism for not exploring important issues beyond the taste-based vs. statistical discrimination debate. Particularly, it has been pointed out that not enough attention has been paid to internalized racism, or self-racism the process by which individuals come to discriminate against members of their own group and the ways in which racist assertions are believed by those who are the targets of discrimination (Hoff and Stiglitz 2010, Bell, Berry, Marquardt and Galvin Green 2013). Our study attempts to identify and distinguish between the different forms of racism by using different information treatments in a laboratory setting.

The paper is organized as follows. In the following section, we describe the experiment design, which was designed to have subjects on both sides of the labor market. In section 3, we provide a brief summary of the subjects participating in the experiment and how session participants were rated externally by a sample from a separate population. Section 4, briefly describes that there were no significant racial differences in regards to performance and grade point average. In section 5, we discuss the observed signaling/lying behavior observed in the stated score treatment. In section 6 we discuss the most interesting results shown by this experiment design, the hiring choices and discrimination that occurs across different treatments. This is followed by our conclusion where we discuss the meaning of these findings and potential future directions based on our findings.

2 Experiment Design

We simulate a job-market in which participants are able to "hire" and "be hired" based on information that is varied across seven distinct rounds. The design of the experiment allows us to measure the extent of inter-group discrimination, internalized discrimination and whether these are statistical or taste-based. The experiment was designed to provide information on both sides of a

labor market: the candidate or labor supplier, and the employer or labor demander. Individuals were paired into teams/firms based on a variety of information treatments that were used to signal the quality of potential candidates as related to other individuals participating in each session. To simulate the job application process, in one of the treatments subjects had the opportunity to state their performance on a previous round of the task. This treatment was designed to provide subjects to lie or inflate their performance to potential employers– this is a practice observed in the evaluation of many resumes by HR professionals where candidates lie or inflate past job experiences or performance (George and Marett 2004).[1]

2.1 Experiment Task

Subjects performed the task in a computer lab on desktop machines with privacy blockers to maintain anonymity regarding their choices. A total of seven two minute rounds were performed where subjects had a two minute time limit to solve as many word anagrams as possible. The anagrams consisted of four to six scrambled letters.[2] Sometimes the letters could create more than one word. For example, the following letters have multiple solutions:

P S A E H

Possible correct solutions:

SHAPE

PHASE

If an answer was provided subjects then moved on to the next anagram and their responses were graded as being correct so long as the word existed in the English language. In each round, task performance was measured as the number of anagrams solved.

After each round subjects were asked a series of questions regarding their beliefs about theirs and other's performance. These questions included the number of anagrams they believed they solved correctly; the group average, the number of anagrams solved by the best person in the group and the number they believed their partner/teammate solved correctly. To ensure that subjects had an incentive to answer accurately they received the equivalent of the piece rate for each belief question they answered accurately if that round was chosen.

[1] As a recent example see former CEO of Yahoo! (Scott Thompson) who resigned in 2012 for lying on his resume.

[2] The letters came from words that are among the 3,000 most common words used in American magazines and newspapers.

Upon completion of the job market simulation task subjects performed a risk aversion task similar to the one used in Holt and Laury (2002). This was done in part to examine whether risk aversion plays a role in the inflation of stated performance. This was followed with an exit survey that asked a variety of demographic questions as well as some subjective belief questions.

2.2 Payments and Treatments

The first round was a a practice round for which subjects were to get acquainted with the task. After the practice round subjects were informed that six rounds would follow and that one of the six rounds would be randomly chosen to be used for payment. The second round paid subjects a piece rate of \$0.5 for each anagram solved. The other rounds involved selecting teammates and team competitions for which subjects were paid for performance on a relative scale and three of the selection and competition rounds were randomized. The rounds were as follows:

1. Practice with no payment.
2. Piece-rate: paid a piece rate for each anagram solved.
3. (Randomized) Team competition: choose a teammate based on picture.
4. (Randomized) Team competition: choose a teammate based on a list of Round 2 performance.
5. (Randomized) Team competition: choose a teammate based on stated score.
6. Team competition: choose a teammate based on picture and stated score.
7. Team competition: choose a teammate based on picture and real performance.

After each round subjects were asked a series of belief questions for which they were paid for if they were accurate. Subjects also received feedback about their own performance after the practice round and the piece rate round. For the team competition rounds feedback was not provided until the end of the experiment session. This was to ensure that knowledge about the team's performance did not lead to individuals altering choice behavior in subsequent rounds.

2.2.1 Payouts: Selection and Team/Firm Competition

The team competitions involved two parts for which subjects were paid. First, each participant was rewarded for being chosen as a teammate by someone else in the session by receiving a flat rate of \$4 for each time they were chosen. Thus, there was a strong incentive to be chosen as a teammate by subjects. This was to simulate the job application process where candidates apply for positions at first by sending out a resume (or CV). Candidates in each round could be chosen by multiple parties, which is akin to getting multiple job interviews.

The payout for being chosen as a candidate is the prize of \$4 multiplied by the number of individuals in the group that choose that candidate. If we define I as the number of times a candidate is selected then $0 \leq I \leq n-1$ where n is the number of session participants. The payout for being selected as a candidate in a session is shown in Equation 2.1 which provides one part of the payout for the experiment.

$$Payout_{select} = I \times \$4 \tag{2.1}$$

Each participant had the opportunity to choose a teammate to pair with for the team (firm) competition. The team competition rewarded the individuals choosing their teammate with a dollar value that was based on relative ranking of their team's performance for that given round within the session. The payout formula is shown in Equation 2.2, which shows that payouts depended on session size (n) and an individual's team rank within the session. As an example, if a session had a total of four individuals participating then the top ranked team would receive \$16, the second would receive \$8, the third would receive \$4 and the fourth ranked team would receive \$2.[3]

$$Payout_{comp} = n \times \$8 \times 0.5^{rank} \tag{2.2}$$

It is important to note that individuals never had the opportunity to choose themselves; consequently it was in each subject's best interest to choose the best possible candidate. The chosen candidate did not receive any payouts from the team beyond the flat rate of \$4 for being chosen. In theory, if all subjects have accurate information concerning potential candidates' abilities then they would all choose the best candidate as a teammate and the team competitions would depend on just each individual's performance while at the same time one candidate would be handsomely rewarded for being chosen multiple times by all the possible teams.

The experiment was designed to provide subjects with incentives towards two goals. Each subject has an incentive to be chosen as a teammate by as many other subjects as possible, the flat rate payment for being selected provides this incentive. As well each individual has an incentive to choose the best possible candidate for her or his team which would provide them with the best chance in getting a higher ranking in the team versus team competitions that occur in each of these team competition rounds. Thus, the total payout for the labor market portion of the experiment consists of one randomly chosen round and is the sum of $Payout_{select}$ and $Payout_{comp}$.

[3] In the event of a tie a random draw determined the ranking among the tied teams.

In selecting candidates, three distinct randomized (by session) information treatments were used: a picture, real performance from the piece rate round, and stated performance which can be thought of as a lie or noisy signal of quality. These three random treatments were followed by two other information treatments: the next treatment used the stated performance posted next to the candidate's photograph, which was then followed by the candidate's real score posted next to the candidate's photograph.[4] Thus, any signaling in the stated score round was limited to one round to limit the strategic set of options. A potential optimal strategy for candidates is to send a signal to be chosen by as many session members as possible and to choose the best possible candidate available. Also, given that the team competition was not a winner-take-all competition, there is an incentive to always solve as many anagrams as possible to improve a subject's relative ranking.

2.2.2 Information Treatments

Picture Treatment: Prior to the experiment session, portrait photographs were taken of subjects. These pictures were used to make racial differences salient without explicitly stating that the picture was designed to explore racial discrimination. In one of the randomized treatments, photographs of all the individuals (other than the subject choosing) in the session were presented on subjects' computer screens. Subjects then selected one picture for that individual to be part of their team for that round. In this picture treatment no information was provided regarding actual or stated scores.

Real Performance Treatment: In the real performance treatment a list of actual scores was provided on each subject's computer screen with a button to make a selection. These real scores came from the piece rate round and all subjects were informed as to the source of these performance scores.

Stated Performance Treatment: Subjects were informed that they get to state their performance and that other subjects will choose teammates based on their stated scores. Subjects were also informed that they will be picking partners based on others' stated scores.[5]

There were seven total rounds in each experiment session where rounds three to five were information treatments that were provided in random order for different sessions. This randomization is necessary because one of the treatments allows subjects to state their performance from the piece rate round as opposed to just having it reported accurately and automatically by the soft-

[4]Real scores and stated scores with photographs were the same as were used in previous rounds so as not to allow candidates to alter behavior or signals.

[5]See a copy of experiment instructions in Appendices.

ware application. Thus, some subjects had the opportunity to see everyone's real performance in the session prior to the stated score round, they had better information about the distribution of abilities than individuals who did not see real performance before the stated score round. Given that there is no truth-telling incentive or constraint in this experiment, individuals see little to no cost from lying in regards to their statements regarding past performance.

Picture + Stated and Picture + Real Treatments: The final two rounds were designed to test whether choices made in the earlier treatments remained consistent when stated performance and real performance information was provided with the pictures of the individuals. These final treatments were not randomized given that individuals could identify actual performance if the picture plus real performance was ever shown first. Therefore, the last round to be shown was the picture along with the real performance. The real performance was based on the original piece rate round and the stated performance came from the round where individuals stated their performance. This comparison of stated and real performance with the pictures of subjects provides us with the ability to compare the consistency of teammate/partner choices when multiple pieces of information are provided to subjects as part of their screening process.

3 Subjects and Sessions

Subjects were paid a $5 payment for participating in the experiment and earnings they made from the experiment were then added to this participation payment. Along with the Job Market/Team Competition experiment, subjects were also paid for a risk aversion. The experiment took between fifty to seventy minutes to complete. A total of 115 subjects participated in the study and on average subjects were paid $24.01.

A total of nineteen sessions were run with session sizes ranging from three to nine subjects with an average session size of six subjects.[6] Given that the experiment was focused on race, the racial composition of subjects session groups is important. Subjects were from a public state university in the Mid-West United States with a student population of approximately 23,400 students. The university population the sample was drawn from is approximately 63% White and 57% female.[7] In using (not mutually exclusive) self-reports for race the we find the subject pool was 64% White,

[6] We control for session size in payouts and analyses and find the race effect remains consistent.

[7] Population data is from the National Center for Education Statistics (NCES) for the approximate time the study took place (2011-2012).

27% Black, and 12% Asian. Sessions were racially diverse as only one session contained only White subjects and one session contained all non-White subjects.[8]

Table 1 provides the summary statistics of the subjects who participated in the study. The subjects are representative of the university population which has a larger percentage of Black individuals than the national average of 12% for public 4-year institutions in the United States (National Center for Education Statistics, Fall 2013). Demographic data was collected using a self-reported survey which took place in sessions after the experiment tasks were complete. One question asked whether subjects had previously felt disadvantaged because of their race: we consider this to be self-reported discrimination.[9] Table 1 shows that 25% of subjects believed that they had previously been discriminated against. Thus, not only is the set of subjects in the study fairly racially diverse, but a significant portion of them believe they had experienced racial discrimination in the past.

[Table 1 about here.]

In Section 2 it was explained that in some of the treatments subjects chose partners based on photographs presented to them of other people in the session. To measure physical attributes that may impact the choice of partners, the pictures were rated by an external group of individuals using Amazon Mechanical Turk. Each individual was rated on a whole number scale between one and seven by fifty-four individuals across a number of attributes including: attractiveness, generosity, honesty, and intellectual ability. The goal of these ratings was to provide a measure of physical characteristics which may influence subjects' selections of partners. The values used as control variables were averaged across all the individuals who rated each subject's picture. Even by averaging we see from Table 1 that there was a great deal of heterogeneity between subjects across all the characteristics that were rated.

Along with physical attributes that are observed in the photographs, the possibility exists that subjects may have known each other from previous interactions. Consequently, subjects who are close to each other socially may choose each other due to the potential payment each individual receives if they are chosen. To control for this form of favoritism, subjects were asked to rate (on a one to seven scale) photographs of each other upon completion of the experiment– after they

[8]The probability of these same race sessions occurring based on population statistics was 7% and 5% respectively. Thus, a priori we would expect that one session of each would occur based on the total number of sessions and session sizes used for the study.

[9]The wording was: *have you experienced disadvantages in the last two years because of your race?*

had been assured of their payout amounts.[10] A rating of seven meant that a subject considered themselves as socially distant from each other while a rating of one meant that subjects knew each other very well. The Closeness statistic summarizes these social distance measures. We find that most subjects did not know each other well, because 88% of closeness measures were 7. Approximately 95% of subjects rated each other as 5 or higher, which suggests that few subjects knew others really well, but given the potential influence of previous social ties we are able to control for any favoritism based on subjects previous social proximity using this closeness measure.

4 No Race Differences: Performance and GPA

The word anagram task was fairly race neutral in terms of performance. Blacks had a mean performance score of 9.24 and a median score of 9.00 across all rounds. Non-Black participants had a mean score of 9.55 and a median score of 9.00. Learning did occur across the rounds but we control for these learning effects in regressions in Table 2.

In Table 2 we use random effects panel regressions to estimate performance scores while controlling for a number of other factors including age, race, gender, and learning. We find that GPA and learning are significant predictors of performance. We also find that risk aversion is negatively correlated with performance.[11] Other than the sample of Asian subjects (12% of subjects), we find little evidence of differences in performance based on gender or race.

[Table 2 about here.]

We also find little difference between the grade point average (GPA) of subjects based on race. Table 3 shows that the GPA of Blacks is lower, but we find the difference is not statistically significant (p-value=.227) between subjects participating in the this study. Given the insignificant differences based on race regarding performance and GPA there is little incentive to discriminate based on race when choosing teammates; in other words, there should be no statistical discrimination based on race for this task.

[Table 3 about here.]

[10] The wording was: *It is possible that you have had some personal interactions with some of the people in this session. To take this into account, the researchers would like you to indicate your social closeness with each of the individuals in the pictures below. A rating of 1 means you are very close to the individual. A rating of 7 means you are very distant from that individual.*

[11] We believe this may occur because those subjects are less likely to guess.

5 Signaling or Lies

When given an opportunity to lie about one's score, optimizing behavior would dictate an inflation (or potentially deflation) of reported scores– a lie akin to the "padding" of a resume.[12] We observe that lying in general is amplified with knowledge of information about the abilities of competitors. A potential reason for this behavior is that participants may not lie until they have an idea of their actual performance relative to others. Then the stated/lie score falls within the range of a believable signal. Based on our examination of the signals candidates send, we find that without information regarding the distribution, subjects send a signal not significantly different from their beliefs of the best score. Once the distribution of scores is known individuals send signals similar to the known best scores.

5.0.3 Observed Strategies for Stated Scores

Before stating scores, each subject formed a prior belief about the distribution of scores. If the real scores have been shown prior to having the opportunity to provide a stated (lie) score then all subjects have the same information set and similar beliefs about the distribution of scores. We assume that subjects desire to be chosen due to the payoff of being selected. Therefore, they should send a signal that leads to them being chosen as frequently as possible. For simplicity, we assume that subjects who send an equal and optimal signal are chosen randomly. In this one-round signal sending game a potential strategy is to send a signal that makes an individual look like the best performer in the group without seeming like an outlier which would make the signal seem unbelievable.

To consider behavior in this stated score round, assume that individuals all have the same information set concerning everyone's performance (assume they see the list of real scores first). Then a believable signal can only consist of a signal that is less than or equal to the highest real score ($\overline{r}$) and greater than or equal to the lowest score ($\underline{r}$). Then individual i must choose a signal $s_i \in [\underline{r}, \overline{r}]$. Given that there are no repeated rounds of signaling and no verification it is a likely strategy to send a signal $s_i = \overline{r}$. There are two possible situations regarding signaling behavior in the stated score round. First, all other candidates (j such that $j \neq i$) send a signal $s_j = \overline{r}$, then the probability that an individual gets picked is random with a probability equal to $\frac{1}{n-1}$. Given

[12] Feltovich et al. (2002) show that deflation of scores could be reasonable in certain situations.

that $n-1$ individuals are picking, this leads to an expected payoff of being chosen once and each individual receives an expected value of \$4. A second possibility is that at least one individual j sends signal $s_j \neq \overline{r}$. Then it maybe the case that people that choose candidates disregard the signal because it is less than or greater than $\overline{r}$ and all individuals that sent signals equal to $\overline{r}$ have an improved probability of being chosen. Or it may be the case that individuals believe the candidate is sophisticated and is sending a signal to differentiate themselves from others. For this strategy to be worthwhile there must be more than one individual that would choose the candidate or else the individual would be no better off than sending a signal equal to $\overline{r}$.[13]

There are numerous possibilities regarding signals being sent and the support depends on beliefs. The reason we use this format of signal sending without verification is because it is found that resume lies and inflation are fairly common (George and Marett 2004). Often lies, omissions, inflated statements go unnoticed or may even take years to identify. In the experiment, participants have an incentive to lie by inflating their scores to a degree by which they might be chosen as a partner, but not to such a degree that others would suspect them of lying– leading to a choice not to select the individual in question.

Awareness of other people's scores by having the real score round come before the stated score round, was revealed to be important. On average, participants were found to lie more once they were aware of the real scores of others in the group. Thus, access to information about the scores of competitors caused more deception. On average the stated score was 0.13 higher than the observed best score, which was not significantly different from zero. For individuals who did not know the best score, we compared individual's stated scores to their beliefs of the best score in the previous round. We found that these individuals stated scores were on average 0.46 less than their beliefs of the best score and this is was not significantly different than zero.

A number of factors affect lying to "employers." In our experiment, we still find a certain degree of variability away from the mean behavior. Table 4 provides linear regression estimates for the stated (Lie) scores of individuals. The Table shows that Black subjects, especially Black males, are more likely to state lower scores than others. This remains true even when controlling for the Real Score. The impact of real scores (from the real score round) is significant.[14] Once we include

[13] If everyone believes that a signal not equal to $\overline{r}$ suggests sophistication and an actual signal of quality is disregarded then a babbling equilibrium is possible (Sobel 2009). But based on observation, in our experiment these prior beliefs seem unlikely to support this behavior.

[14] The significance of lying that takes place by Asian subjects seems to be mainly caused by them performing worse than others.

the beliefs of performance from the previous round then observing the real performance round is no longer significant. Thus, individuals are updating their beliefs about performance and therefore lying based on the most recent update as opposed to using the real performance from the piece rate treatment.

[Table 4 about here.]

We find the Black subjects (especially males) tend to inflate scores less than others. Overall we find that individuals lie or state scores that are based in part on their recent beliefs of their own performance. This self belief is more important than other peoples scores or even actual scores in shaping stated scores. But based on stated scores, if information was publicly available we would expect no statistical discrimination when subjects stated scores with race being salient.

6 Hiring Choices and Discrimination

In each of the team competition rounds each subject has an opportunity to choose/hire a teammate/partner. We first examine the proportions chosen across a number of obvious characteristics to examine if there are any general differences in choice behavior that coincide with the different treatments. Table 5 has the proportion of subjects chosen according to the following characteristics (in order of columns) Female, Black, White, Previously discriminated against because of race, Real Score (from Round 2), Stated Score (from stated score treatment). Each row in the table is for the labeled information treatment.

From Table 5 we observe that the proportion of individuals chosen by race seems to vary a great deal based on the treatment. We observe that Black individuals are chosen less when a picture is shown. Alternatively, the proportion of White subjects chosen is greatest during treatments that show a picture and the proportion is highest when the photograph is shown with the real scores. Subjects who have previously been discriminated against also are chosen less when their pictures are part of the information set.[15] Thus, the treatments exhibit some variation based on the choices of individuals chosen/hired, but this variation seems to be based on race.

We would expect that treatments when real scores are shown would lead to those treatments having the highest average real scores. Table 5 shows that average real scores are highest when

[15] We find that there is no difference in the attractiveness of subjects that have been discriminated against and others.

real scores are shown with pictures of individuals (13.79) and when just real scores are shown the mean is slightly lower at 13.10.[16] But both those treatments have significantly higher real scores of individuals chosen than any of the other treatments. Meanwhile, both the single information treatments of the picture and the stated score lead to selecting lower quality candidates. The final column shows that the highest average stated score actually occurs when the picture is shown along with the stated score and not when the stated score is the only information available for making a choice. This difference suggests that subjects are not always choosing the highest stated score that is presented to them. But the table suggests that real scores are positively related with the stated scores as the means of real scores are high when the mean stated scores are high as well.

[Table 5 about here.]

In the picture round, the only information available consisted of the facial characteristics of individuals. Subjects had to form beliefs of the abilities of these candidates from a very unrelated source of information, but when making choices some logic must follow in making decisions.[17] From a picture one can make subjective judgments on a person's attractiveness, intelligence and trustworthiness. To test whether subjects were choosing partners from the picture treatment based on these types of judgments each photograph was rated externally across a number of characteristics. We used external judgments to get a larger number of judgments on each individual, to limit experiment session length and also to not have subjects feel judged within a live experiment session, which could make them uneasy. Table 6 provides the average external ratings (on a one to seven scale) of subjects chosen along with the mean GPA of chosen subjects. The characteristics that the external judges were asked to measure include (in order of table columns) attractiveness, generosity, honesty and intelligence. The table shows that all these dimensions are highest when individuals had only the photographs to use to choose candidates. We also find that these subjective measures of the photographs do not correlate with a measure of intellectual ability such as GPA.

[Table 6 about here.]

[16] A potential explanation for the real score average not being highest when that was the only information available is that if a high score was an outliers it was not believed in earlier treatments.

[17] Eckel and Petrie (2011) find subjects value pictures prior to making choices in trust games.

6.1 Testing for Discrimination

To test for racial discrimination, we examine the probability an individual is hired based on the characteristics available in each treatment. To examine whether discrimination is statistical or taste based we compare the hiring (partner selection) behavior of White subjects versus Non-White subjects. We assume that if discrimination is statistical then we should observe discrimination in the same direction by both groups. By having independent treatments where only a single piece of information is available we are able to compare changes based on race and information context once real scores and stated scores are combined with the photographs. We estimate probabilities by estimating the marginal effects from probit regressions where the dependent variable is whether a candidate was chosen/hired or not.

Table 7 compares the probabilities of individuals being hired in just the Picture treatment broken down by the race of the individual who is hiring. The regressions are separated for individuals hiring if they are White versus if they are Non-White.[18] The size of sessions is also controlled for, but the general, and intuitive result is that the larger the number of candidates in the session the less likely an individual will be hired. When the individual making the choice was White we observe that an individual is more likely to be hired/selected if they look to be more intelligent in the photograph and if they are socially close (the closest an individual can be to someone is 1 and the furthest they can be from someone is 7) to the individual that is hiring. Closeness is the only variable (not including controls) that is significant in the same direction for the Non-White individuals selecting partner.

In terms of discrimination, we find that if a White individual is hiring then a subject is more likely to be hired if there is a larger proportion of Black individuals in the candidate pool: so long as the candidate themselves is not Black. There is approximately a .12 lower probability of being selected by a White individual if a subject is Black. We observe no such discrimination by the Non-White population. Based on the picture treatment alone we observe racial discrimination by White subjects when choosing a partner.

[Table 7 about here.]

Table 8, focuses on the Stated Score treatment. In this treatment individuals choose a score to report as their performance to be posted alongside others from which people would potentially

[18] We get qualitatively equivalent results when classify hiring individuals as Black and Non-Black. White was chosen on the hiring side to provide a more homogeneous racial population for the majority.

choose them as a candidate. The table breaks down the probabilities of being hired by whether the individual doing the hiring is White or Non-White.[19] We find that these two groups respond to the stated scores differently.

White individuals are more likely to select higher stated scores, but do so at a decreasing rate as exhibited by the negative coefficient of the square of the stated score. Otherwise, they respond to very little of the other information available such as the top stated score. Interestingly, if the treatment using real scores was shown previously it has little impact on the selection of candidates by White subjects. The same cannot be said of the Non-White subjects: when making selections an individual is less likely to be hired if there is a high "Top Stated" score that does not belong to them. Also, if a high real score was shown previously it lowers the probability of being selected by a Non-White subject by at most 0.1. Without pictures we find that there is no discrimination by either group in this treatment as no information about race is provided.

[Table 8 about here.]

Table 9 provides the probabilities from the Real Score treatment where subjects could choose/hire partners from a list of real scores. The Table shows that both White and Non-White subjects made candidate selections in a very similar manner. A candidate has a higher probability of being selected the higher their real score is and a lower probability of being chosen the higher the top score is in the session.

[Table 9 about here.]

The three separate single information treatments were all followed by the Stated Score and Picture treatment. Each individual's previously stated score was listed next to their picture. In this treatment we find that both White subjects and Non-White subjects discriminate against hiring/selecting Black candidates. Figure 1 provides the estimated marginal effects for each group based on the Stated (Lie) Score with the shaded region showing the standard errors across the estimates. The figures show that a Black candidate is less likely to be chosen when they can be identified as Black and they report a stated score.

[Figure 1 about here.]

Table 10 shows the marginal effects at the median for three different probit estimations. The first column provides marginal effects for the entire subject pool, which shows that the stated score

[19] Using Black versus Non-Black provides us with qualitatively equivalent results.

has an impact on selections as does the social closeness of the subjects with the individual choosing partners. But the most striking result across all three estimations is that a Black candidate is less likely to be selected regardless of the race of the individual doing the hiring. Interestingly, the two groups respond differently to photographs specifically using the external measure of intelligence: we find that White subjects are more likely to select individuals who look intelligent while the Non-White subjects are less likely to do so. Overall, the result that both White and Non-White subjects are less likely to choose Black candidates suggests that this type of discrimination is statistical. Specifically the discrimination seems to be tied to a lack of trust towards Black subjects when they can lie/inflate their reported performance. We discuss this possibility in our conclusion.

[Table 10 about here.]

In the final round, subjects choose partners based on real scores from the piece rate round which were listed next to the pictures of individuals. We find support for taste based discrimination by White subjects in this round. In general, we find that the real score plays a major factor in choosing a partner from potential candidates. Figure 2 provides the cumulative probability of an individual being hired based on real scores and it shows that discrimination against Black candidates is only performed by White subjects. Table 11, which provides the marginal effects of a candidate being hired at the median shows that discrimination is entirely performed by White subjects though it shows up as significant in the estimations for the entire subject pool. This result suggests that White subjects are using taste-based discrimination against Black candidates even when given accurate performance information along with individual pictures. On the other hand, subjects that are Non-White mainly choose candidates based on performance and social proximity–the closer the individual, the more likely they are to be picked.

[Figure 2 about here.]

[Table 11 about here.]

7 Conclusion

Our experiment design allows us to distinguish important features relating to race and discrimination. We can discern the extent to which racism exists in our study based on different informational contexts. This is achieved by comparing individual team selection choices based on changing information: with pictures only, past scores in a similar task only, and a combination of the two.

This allows us also to determine the extent to which racial discrimination exists. We also observe the impact of potential lies in quality signals by race categories and consequently observe a bias in trusting signals from Blacks when selecting teammates. If this lack of trust of Black signals was based on statistical discrimination we would expect it to disappear in both groups when controlling for real performance scores when included with the picture. But this does not happen amongst white subjects who continue to discriminate even when actual performance is available– providing evidence of taste-based discrimination.

We find that internalized discrimination of Blacks is statistical and is based on a noisy signal (incomplete information). We also find racism of Whites against Blacks is taste-based and persistent across information treatments whenever race is salient. Some of this discrimination stems from a lack of trust in Black signals. But Blacks lie less than Whites on average and both Non-Whites and Whites expect Blacks to lie more than Whites in this study. The finding that Non-White subjects are discriminating against Blacks is likely based on assumptions or prior beliefs and seems to indicate a lack of trust of Black subjects even among themselves. In this case, the statistical discrimination was inaccurate, but likely culturally and socially motivated. Information, in this case, is a salve for discrimination and therefore a dissemination of non-race information should diminish the extent to which this incorrectly formed statistical discrimination occurs.

Some of this bias against Blacks may be the result of race-based educational achievement gaps, which remain pronounced across the globe, including in the United States(Singleton 2014, Carnoy and Rothstein 2013). The enactment and funding of comprehensive educational attainment programs– many of which have been advocated for by educational researchers– would likely do much to diminish the statistical gap on which intra-group Black discrimination occurs. But without changing the gap, this type of discrimination can be seen as being efficient in many respects and may even be based on the racial performance gap shown in SAT scores.

The largest of the problems associated with discrimination is taste-based discrimination emanating from the Whites in this study. Since taste-based discrimination is not informational, there are limited policy choices available for counteracting this. One of the most obvious choices in this regard would be the implementation of affirmative action policies requiring the hiring of a minimum number of Black applicants as long as they are equally qualified as White applicants. Taste-based discrimination is economically inefficient and in the absence of perfectly competitive markets, which would compel discriminators to change their behaviors via threat of competition.

Such affirmative action programs could, if designed correctly, improve profitability for businesses whose hiring practices are controlled by Whites. Another solution for taste-based discrimination is the transformation of these biases and/or tastes. But given the inability of good information about abilities to remove taste-based discrimination, it suggests that these preferences are unfortunately deeply rooted, therefore providing a valid case for the implementation of affirmative action programs.

The findings regarding lying behaviors have important implications, especially regarding the area of community political and economic development. It has long been understood that trust is a primal component of social capital. And social capital, especially in communities, is essential for smoothly running economies, political systems, and civil society groups(Putnam 2013). The finding that Blacks tend to lie less in general can be thought of as a social capital asset. This asset is most pronounced when Blacks are aware that others know their identities. The finding that Blacks expect other Blacks to lie more, on the other hand, implies an erosion of social capital via internalized discrimination. Finding the cause of this internalized bias has value because there is value within the group to know who your neighbors are and currently, among this subject pool, Black subjects are unnecessarily considered less trustworthy by other Black subjects.

The unique design of this study allowed us to identify both taste-based and statistical discrimination amongst various groups. We find that Whites discriminate against Blacks regardless of information suggesting taste-based discrimination and find that Non-Whites statistically discriminate against Blacks based on what seem to be inaccurate beliefs about honesty. In general, the forms of discrimination we find suggest a very serious problem. Not only is the issue of taste-based discrimination a problem, but the fact that in-group discrimination has incorrectly manifested itself in the controlled setting of the lab suggests that minority groups may be shouldering unnecessary biases onto one another. Better information about actual abilities seems to be the best way to make these biases irrelevant among discriminated groups.

References

Apel, D.: 2014, Hands Up, Don't Shoot: Surrendering to Liberal Illusions, *Theory & Event* **17**(3).

Arrow, K.: 1973, The theory of discrimination, *Discrimination in labor markets* **3**(10), 3–33.

Becker, G. S.: 2010, *The economics of discrimination*, University of Chicago press.

Bell, M. P., Berry, D. P., Marquardt, D. J. and Galvin Green, T.: 2013, Introducing discriminatory job loss: antecedents, consequences, and complexities, *Journal of Managerial Psychology* **28**(6), 584–605.

Bergmann, B. R.: 1971, The effect on white incomes of discrimination in employment, *The Journal of Political Economy* pp. 294–313.

Bertrand, M., Chugh, D. and Mullainathan, S.: 2005, Implicit discrimination, *American Economic Review* pp. 94–98.

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

Card, D. and Rothstein, J.: 2007, Racial segregation and the black–white test score gap, *Journal of Public Economics* **91**(11), 2158–2184.

Carnoy, M. and Rothstein, R.: 2013, What do international tests really show about US student performance, *Retrieved from Economic Policy Institute website: http://www. epi. org/publication/us-student-performance-testing* .

Castillo, M. and Petrie, R.: 2010, Discrimination in the lab: Does information trump appearance?, *Games and Economic Behavior* **68**(1), 50–59.

Castillo, M., Petrie, R. and Torero, M.: 2012, Beautiful or White? Discrimination in group formation, *GMU Working Paper in Economics* .

Eckel, C. C. and Petrie, R.: 2011, Face value, *The American Economic Review* pp. 1497–1513.

Edgeworth, F. Y.: 1922, Equal pay to men and women for equal work, *The Economic Journal* pp. 431–457.

Fawcett, M. G.: 1892, Mr. Sidney Webb's Article on women's wages, *The Economic Journal* **2**(5), 173–176.

Feltovich, N., Harbaugh, R. and To, T.: 2002, Too Cool for School? Signalling and Countersignalling, *RAND Journal of Economics* **33**(4), 630–649.

Fleming, J.: 2002, Who will succeed in college? When the SAT predicts Black students' performance, *The Review of Higher Education* **25**(3), 281–296.

Fletcher, J. M. and Tienda, M.: 2015, High school quality and race differences in college achievement, *Race and Social Problems*, Springer, pp. 137–159.

George, J. and Marett, K.: 2004, The truth about lies., *HR Magazine* **49**, 87–91.

Gneezy, U., List, J. and Price, M. K.: 2012, Toward an understanding of why people discriminate: Evidence from a series of natural field experiments, *Technical report*, National Bureau of Economic Research.

Hoff, K. and Stiglitz, J. E.: 2010, Equilibrium fictions: A cognitive approach to societal rigidity, *Technical report*, National Bureau of Economic Research.

Holt, C. and Laury, S.: 2002, Risk Aversion and Incentive Effects, *American Economic Review* **92**(5), 1644–1655.

Labor Force Characteristics by Race and Ethnicity: 2014, *Technical Report 1050*, U.S. Bureau of Labor Statistics.

Men, B. W. O. B.: n.d., Black Student College Graduation Rates Remain Low, But Modest Progress Begins to Show, *Women* **45**(34), 40.

Putnam, R.: 2013, The prosperous community: social capital and public life, *The American Prospect* **13**.

Riach, P. A. and Rich, J.: 2004, Deceptive field experiments of discrimination: Are they ethical?, *Kyklos* **57**(3), 457–470.

Riach, P. A. and Rich, J.: 2006, An experimental investigation of sexual discrimination in hiring in the English labor market, *Advances in Economic Analysis & Policy* **5**(2).

Rich, J.: 2014, What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000.

Rubinstein, Y. and Brenner, D.: 2013, Pride and prejudice: Using ethnic-sounding names and inter-ethnic marriages to identify labour market discrimination, *The Review of Economic Studies* p. rdt031.

Singleton, G. E.: 2014, *Courageous conversations about race: A field guide for achieving equity in schools*, Corwin Press.

Sobel, J.: 2009, Signaling games, *Encyclopedia of Complexity and Systems Science*, Springer, pp. 8125–8139.

Appendices

Appendix A Experiment Instructions

In this experiment you will be performing a simple task for real money. The task will consist of solving as many word anagrams as possible within X minutes. Before you begin a real session, you will get one practice round so that you will understand how the task works.

The word anagrams will consist of 4 to 6 scrambled letters that can make a word that is among the 3000 most common words in used in American newspapers or magazines. Sometimes the letters can create more than one word. For example, the following letters:

P S A E H

This anagram can be solved correctly by submitting the word SHAPE or the word PHASE. Your answer will be counted as being correct so long as you form an English word with the scrambled letters. After you submit one word another set of letters will be shown to you. Your goal is to solve as many anagrams as possible within the X minute time limit. You will not be penalized for guessing or submitting incorrect words, but you will only be paid for correct words.

After the practice round, you will perform the task in six other rounds. One of the rounds will be chosen for your experiment payment. Your payout will in part depend on how well you perform in the task compared to others.

One of the six rounds will be used as part of your payout for the experiment. Please continue to the practice round.

For RID 2 (Piece-Rate):

In this Round you will be paid Y dollars for each anagram you solve in X minutes if this round is chosen as part of your experiment payout.

Your performance in this round may affect payouts in future rounds as other session members may choose to have you on their team based on your performance. You will get a payment of Q dollars in those rounds for each time you are chosen to be part of a team.

If this round is chosen as part of your experiment payout, then you will receive Y dollars multiplied by the number of anagrams you solve in X minutes.

Press continue to begin the round.

In the following rounds you will choose an individual in the session to be part of your group. Your group will be competing against everyone elses group. Your group performance will consist of your performance plus the performance of the individual you are paired with.

Your group performance will be compared to everyone elses group in this session. If this round is chosen as part of your experiment payout then you the payouts based on group placement will be as follows. 1st: $ 2nd: $ In the event of tie, the placement of tied groups will be determined randomly.

Along with this group payment, you will also receive a payment of $ for each time you are chosen to be part of someone elses group. The more times you are chosen, the more money you can make. So if you are chosen by three people in this round then you will make another $. Therefore, the more people that pick you, the more money you can earn.

For RID (Real Performance):

In this round, individuals in the session will choose partners based on the performance from Round 2 (the piece rate round).

You want to pick your group member wisely to increase your chances of winning the group competition.

Press continue to pick your partner.

For RID (Picture)“

In this round, individuals in the session will choose partners based on the pictures taken of you prior to the experiment session.

You want to pick your group member wisely to increase your chances of winning the group competition.

Press continue to pick your partner.

For RID (Stated Performance)

In this round, you will have the option of stating your performance in this task. Based on your stated performance individuals will chose to have you as part of their team. The more people that pick you as part of their team, the more money you will make. For example, if you are chosen by three people in this round then you will make another $ if this round is chosen for your payout.

You will also choose a partner for your group based on their stated performance. You want to pick your group member wisely to increase your chances of winning the group competition.

Press continue to provide your stated performance.

You are free to state any performance you want. But you should remember that the goal is to be chosen most often by people in the session. The more people that pick you as part of their team, the more money you will make, $ for each time you are picked. If you are chosen by three people in this round then you will make another $ if this round is chosen for your payout.

Submit your performance. Choose a partner from the list of stated performances.

For RID 6 (Stated Performance and Picture):

In this round you will choose a partner based on stated performance from the earlier round and their picture. Likewise, individuals in the session will have the opportunity to choose you as a group member based on your stated performance along with your picture.

If this round is chosen as part of your experiment payout then you will receive the group competition payout along with $ for each time you are chosen by someone else.

You want to pick your group member wisely to increase your chances of winning the group competition.

Continue to choose a partner from the list of session members. Choose a partner from the list of session members

For RID 7 (Actual Performance and Picture): In this round you will choose a partner based on their actual performance from the second round and their picture. Likewise, individuals in the session will have the opportunity to choose you as a group member based on your actual performance from round 2 along with your picture.

If this round is chosen as part of your experiment payout then you will receive the group competition payout along with $ for each time you are chosen by someone else.

You want to pick your group member wisely to increase your chances of winning the group competition.

Continue to choose a partner from the list of session members. Choose a partner from the list of session members.

Table 1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Female	0.54	0.5	0	1
White	0.64	0.48	0	1
Black	0.27	0.45	0	1
GPA	2.73	1.37	0	4
Age	25.33	7.7	18	56
Race Discrim.*	0.25	0.44	0	1
n		115		
External Ratings of Pictures (scale 1-7)				
Attractiveness	4.2	0.73	2.32	5.71
Generosity	4.43	0.61	2.94	5.43
Honesty	4.55	0.62	2.78	5.54
Intelligence	4.75	0.52	3.31	5.86
No. of Ratings	53.85	1.93	49	58
Closeness (n=611)	6.68	1.07	1	7

*=Self-report survey question on racial discrimination.

Table 2: **Performance**

	(1) Score	(2) Score
GPA	0.89***	0.86***
	(0.295)	(0.299)
Asian	-5.02***	-4.92***
	(1.222)	(1.239)
Black	-0.29	-0.81
	(0.887)	(1.303)
Female	0.66	0.41
	(0.806)	(0.930)
Black Fem.		0.98
		(1.789)
Constant	5.98***	6.21***
	(1.934)	(1.983)
Age, Learning and Risk Controls	Yes	Yes
Observations	805	805
Number of ID	115	115
r2	0.174	0.176

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: **GPA Differences by Black**

Group	Obs	GPA Mean
Not-Black	84	2.82
Black	31	2.47

Table 4: **Stated (Lie) Score**

	(1)	(2)	(3)	(4)	(5)	(6)
	Lie Score	Lie Score	Lie Score	Lie Score	Lie Score	Lie Score
Female	0.13	0.24	0.10	0.07	0.70	0.51
	(1.566)	(1.516)	(1.034)	(0.956)	(1.106)	(1.059)
Asian	-6.05***	-6.14***	-2.16	-1.50	-1.71	-1.50
	(1.594)	(1.548)	(1.403)	(1.785)	(1.411)	(1.586)
Black	-3.76*	-5.66***	-3.24**	-3.00**	-3.21**	-3.10**
	(1.816)	(1.881)	(1.505)	(1.336)	(1.439)	(1.373)
Black x Female	7.52	7.92	6.25	7.45	7.00	7.40
	(4.658)	(4.571)	(4.720)	(4.627)	(4.337)	(4.450)
Pic. prior	-0.44	0.17	0.58	0.37	0.47	0.39
	(1.592)	(1.815)	(1.239)	(1.264)	(1.259)	(1.266)
Real prior	5.00**	3.28	2.88*	1.48	1.64	1.28
	(1.857)	(2.366)	(1.535)	(1.634)	(1.549)	(1.606)
Prop. Black		8.36*	6.61*	5.67	6.86*	6.29*
		(4.232)	(3.361)	(3.442)	(3.317)	(3.296)
Real Score			0.75***	0.50***	0.26	0.28
			(0.101)	(0.167)	(0.235)	(0.222)
Prv. Belief Top				0.56**		0.29
				(0.225)		(0.182)
Prv. Belief Self					0.89***	0.62**
					(0.299)	(0.223)
Constant	1.24	-3.48	-9.73*	-13.28**	-12.08**	-13.24**
	(4.975)	(5.246)	(4.866)	(5.412)	(4.503)	(4.948)
Controls	Size	Size	Size	Size	Size	Size
Observations	115	115	115	115	115	115
R-squared	0.294	0.316	0.482	0.546	0.558	0.568
r2_a	0.248	0.264	0.438	0.502	0.515	0.522
F	5.161	5.758	20.54	27.73	27.59	26.93
N_clust	19	19	19	19	19	19

Robust standard errors in parentheses
Errors clustered on session
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Chosen Individuals

Treatment	Fem	Black	White	Race Disc.	Real	Stated
Picture	0.57	0.18	0.79	0.13	9.73	13.17
Real	0.57	0.23	0.76	0.19	13.10	16.53
Stated	0.56	0.26	0.69	0.23	9.25	14.76
Stated + Pic	0.56	0.17	0.78	0.11	11.21	16.79
Real + Pic	0.57	0.19	0.83	0.13	13.79	16.36
Total	0.56	0.21	0.77	0.16	11.42	15.52

Table 6: Chosen External Ratings

Round	GPA	Attract.	Gener.	Honesty.	Intel.
Picture	2.77	4.26	4.59	4.70	4.89
Real	2.77	4.02	4.43	4.54	4.70
Stated	2.92	4.23	4.50	4.62	4.79
Stated + Pic	2.98	4.20	4.46	4.55	4.78
Real + Pic	2.86	4.05	4.33	4.43	4.63
Total	2.86	4.15	4.46	4.57	4.76

Table 7: **Hiring By Picture**
Marginal Effects of Being Hired By Picture Only

	(1)	(2)	(3)	(4)
	Hired by White		Hired by Non-White	
Marg. FX	at Mean	at Median	at Mean	at Median
Attractive.	-0.01	-0.01	-0.03	-0.03
	(0.032)	(0.039)	(0.046)	(0.045)
Intelligence	0.14***	0.17***	0.08	0.08
	(0.048)	(0.062)	(0.072)	(0.069)
Closeness	-0.05***	-0.06***	-0.03**	-0.03**
	(0.017)	(0.021)	(0.015)	(0.013)
Prop. Black	0.17**	0.20**	0.12*	0.11
	(0.066)	(0.091)	(0.069)	(0.078)
Cand. Black	-0.12***	-0.13***	-0.05	-0.04
	(0.042)	(0.050)	(0.071)	(0.066)
Cand. Asian	-0.08*	-0.10	-0.08	-0.08
	(0.047)	(0.059)	(0.085)	(0.081)
Cand. Female	0.03	0.03	-0.02	-0.02
	(0.042)	(0.049)	(0.062)	(0.062)
Controls	Order	Order	Order	Order
Controls	Size	Size	Size	Size
Observations	408	408	203	203
Xmfx_y	0.154	0.213	0.184	0.176
N_clust	74	74	40	40
ll	-172.9	-172.9	-95.38	-95.38
chi2	76.07	76.07	112.3	112.3
r2_p	0.0981	0.0981	0.0533	0.0533

Robust standard errors in parentheses
Errors clustered on Individual (Employer/Selector)
*** p<0.01, ** p<0.05, * p<0.1

Table 8: **Hiring By Stated Score**

Hired By Stated Scores

	(1)	(2)	(3)	(4)
	Hired by White		Hired by Non-White	
Marg. FX	at Mean	at Median	at Mean	at Median
Stated Score	1.89e-02***	2.07e-02***	1.92e-02**	1.62e-02**
	(0.005)	(0.006)	(0.008)	(0.007)
Stated Score^2	-3.14e-04***	-3.43e-04**	-2.03e-04	-1.72e-04
	(0.000)	(0.000)	(0.000)	(0.000)
Top Stated	-8.55e-04	-9.34e-04	-4.70e-03***	-3.97e-03***
	(0.001)	(0.001)	(0.002)	(0.001)
Cand. Black	-2.01e-02	-2.16e-02	4.35e-02	3.89e-02
	(0.048)	(0.051)	(0.066)	(0.063)
Cand. Female	2.39e-02	2.50e-02	-4.75e-02	-4.41e-02
	(0.039)	(0.041)	(0.054)	(0.050)
Real Seen	2.89e-02*	3.29e-02	3.70e-02	3.37e-02
	(0.017)	(0.020)	(0.027)	(0.026)
Real Seen x Top	-6.16e-04	-6.73e-04	-3.04e-03**	-2.57e-03*
	(0.001)	(0.002)	(0.002)	(0.001)
Controls	Order	Order	Order	Order
Controls	Size	Size	Size	Size
Observations	414	414	212	212
Xmfx_y	0.168	0.193	0.172	0.134
N_clust	74	74	41	41
ll	-186.3	-186.3	-96.68	-96.68
chi2	578.2	578.2	266.9	266.9
r2_p	0.0416	0.0416	0.0714	0.0714

Robust standard errors in parentheses

Errors clustered on Individual (Employer/Selector)

*** p<0.01, ** p<0.05, * p<0.1

Table 9: **Hiring By Real Score**

Hired By Real Scores

	(1)	(2)	(3)	(4)
	Hired by White		Hired by Non-White	
Marg. FX	at Mean	at Median	at Mean	at Median
Real Score	0.03***	0.03***	0.03***	0.03***
	(0.004)	(0.005)	(0.007)	(0.010)
Best Public	-0.01***	-0.01***	-0.02***	-0.02***
	(0.003)	(0.003)	(0.005)	(0.008)
Cand. Black	-0.03	-0.03	-0.03	-0.03
	(0.032)	(0.031)	(0.071)	(0.087)
Cand. Female	-0.01	-0.01	0.05	0.06
	(0.035)	(0.037)	(0.056)	(0.065)
Cand. Asian	-0.04	-0.04	-0.11**	-0.13*
	(0.053)	(0.052)	(0.052)	(0.075)
Controls	Order	Order	Order	Order
Controls	Size	Size	Size	Size
Observations	414	414	212	212
Xmfx_y	0.121	0.126	0.153	0.212
N_clust	74	74	41	41
ll	-150.3	-150.3	-88.38	-88.38
chi2	73.79	73.79	127.8	127.8
r2_p	0.227	0.227	0.151	0.151

Robust standard errors in parentheses

Errors clustered on Individual (Employer/Selector)

*** p<0.01, ** p<0.05, * p<0.1

Table 10: **Hiring By Picture + Stated Score**

	(1)	(2)	(3)
	Hired	Hired by White	Hired by Non-White
Stated Score	0.037***	0.043***	0.031***
	(0.006)	(0.008)	(0.008)
Stated Score^2	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
Cand. Black	-0.092***	-0.125***	-0.087***
	(0.029)	(0.041)	(0.031)
Attractive.	-0.004	0.012	-0.026
	(0.026)	(0.038)	(0.035)
Intelligence	0.023	0.105**	-0.092**
	(0.037)	(0.052)	(0.044)
Closeness	-0.029**	-0.007	-0.038***
	(0.012)	(0.020)	(0.010)
Prop. Black	0.055	0.116	0.086
	(0.056)	(0.097)	(0.079)
Controls	Size	Size	Size
Observations	611	408	203
Xmfx_y	0.154	0.186	0.116
N_clust	114	74	40
ll	-247.0	-159.8	-78.83
chi2	153.5	123.0	54.27
r2_p	0.156	0.166	0.218

Robust standard errors in parentheses
Errors clustered on Individual (Employer/Selector)
*** p<0.01, ** p<0.05, * p<0.1

Table 11: **Hiring By Picture + Real Score**

	(1)	(2)	(3)
	Hired	Hired by White	Hired by Non-White
Real Score	0.025***	0.025***	0.023***
	(0.003)	(0.004)	(0.004)
Cand. Black	-0.072***	-0.093***	-0.029
	(0.026)	(0.030)	(0.040)
Attractive.	-0.000	0.006	-0.017
	(0.024)	(0.031)	(0.031)
Intelligence	-0.062*	-0.053	-0.049
	(0.034)	(0.042)	(0.049)
Closeness	-0.023**	-0.004	-0.031***
	(0.009)	(0.017)	(0.008)
Prop. Black	0.088*	0.129*	0.023
	(0.053)	(0.077)	(0.059)
Controls	Size	Size	Size
Observations	611	408	203
Xmfx_y	0.134	0.142	0.0901
N_clust	114	74	40
ll	-226.6	-148.6	-75.57
chi2	87.66	66.40	28.65
r2_p	0.225	0.224	0.250

Robust standard errors in parentheses
Errors clustered on Individual (Employer/Selector)
*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Racism in Stated Score and Picture

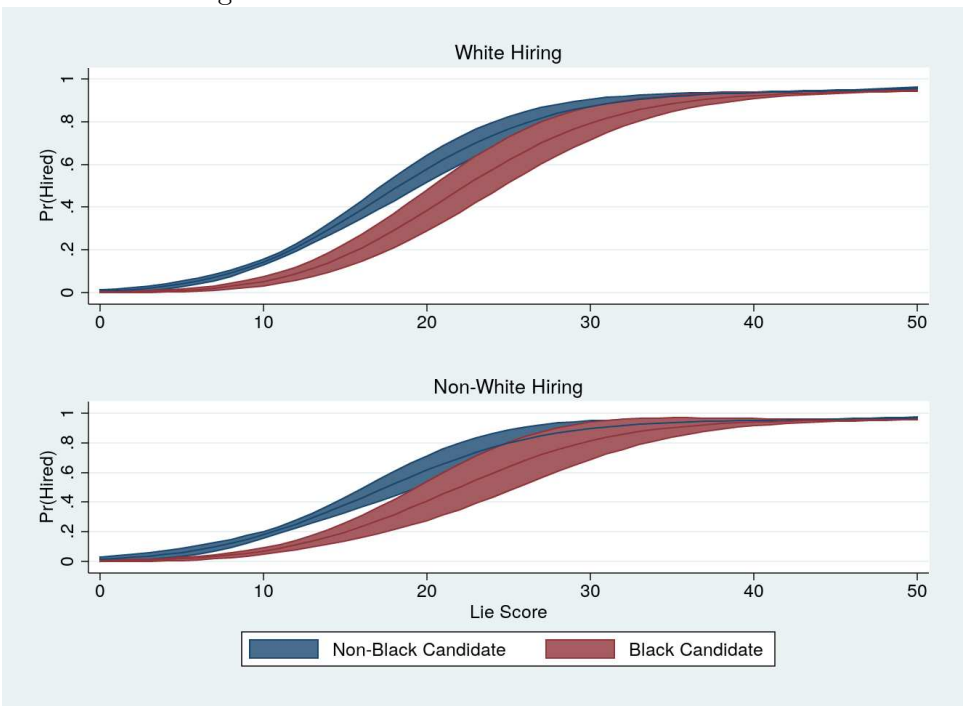

Figure 2: Racism In Real and Picture Round