12-1-2006

Social Barriers to Cooperation: Experiments on the Extent and Nature of Discrimination in Peru

Marco Castillo
Georgia Institute of Technology

Ragan Petrie
Georgia State University

Maximo Torero
International Food Policy Research Institute

Social Barriers to Cooperation: Experiments on the Extent and Nature of Discrimination in Peru

Marco Castillo*
Georgia Institute of Technology
marco.castillo@pubpolicy.gatech.edu

Ragan Petrie
Georgia State University
rpetrie@gsu.edu

Maximo Torero
International Food Policy Research Institute (IFPRI)
m.torero@cgiar.org

December 2006

Abstract
We present a series of experiments to understand the nature and extent of discrimination in urban Lima, Peru. The experiments exploit varying degrees of information on performance and personal characteristics as people sort into groups to test for statistical versus taste-based discrimination. This allows us to examine the nature of discrimination. Our sample is similar to the racial and socio-economic diversity of young adults in urban Lima. This allows us to look at the extent of discrimination. We use a unique method to measure race, along four racial dimensions common in Peru, and find that race is clearly observable. This gives us confidence that we can examine discrimination based on race. While behavior is not correlated with personal, socio-economic or racial characteristics, people do use personal characteristics to sort themselves into groups. Beauty is a robust predictor of being a desirable group member as is being a woman. Being unattractive or looking indigenous makes one less desirable and looking white increases one's desirability. Interestingly, indigenous subjects are three times more likely to be classified as unattractive, suggesting that beauty might mask discrimination. We find that once information on performance is provided, almost all evidence of discrimination is eliminated, except in the most-preferred group. The evidence in these cases is consistent with taste-based, rather than statistical, discrimination.

* The authors would like to thank the Inter-American Development Bank for research funding. We also thank Kevin Ackaramongkolrotn, John List, Andrea Moro, Jorge de la Roca, David Solis, Néstor Valdivia, and seminar participants at the Economic Science Association Meetings and the Southern Economics Association Meetings.

Introduction

There is a growing consensus among Andean countries that racial or ethnic discrimination is not only a grim legacy of colonialism but also an obstacle to fostering democracies and market-friendly economies in the region. For instance, despite the fact that the Peruvian economy grew 77% between 1991 and 2005, half of all Peruvians were still considered poor in 2005.[1] Indeed, little has changed since 1991, when 60% of Peruvians were considered poor (Escobal, Saavedra and Torero, 1998). According to the World Bank's Living Standards Survey in 2000, while the overall poverty rate was 54%, the poverty rate of the population whose mother tongue is Quechua, Aymara or any other native language is 70%. The Peruvian Truth and Reconciliation Commission's final report graphically illustrates the deep divides of Peruvian society: four out of five people killed during the Peruvian civil conflict in the 1980s were indigenous. As reflected by the recent history of ethnic conflict and the emergence of ethnic-based political parties in the region, the issue of exclusion and discrimination has moved to the forefront of political platforms.

While it may be true that the colonial legacy has brought devastating consequences to people of non-European descent, it is also possible that a long history of ethnic conflict has nurtured a sense of distrust among and across different segments of society.[2] This may be important because the lack of awareness and understanding of different ethnic and racial groups could make physical appearance a primary criterion in shaping people's perceptions about each other and determining group association. Such perceptions, based on physical characteristics, may lead people to refrain from interacting with certain segments of the population. Initially, these perceptions might persist even in the face of evidence contradicting them. In the long run, individuals in society may suffer persistent losses due to exclusion if enough sorting takes place.

[1] See Instituto Nacional de Estadistica e Informatica, Lima-Peru, www.inei.gob.pe for detailed information.
[2] This argument is best explained by Walker's (1999) analysis of indigenous uprising prior to the wars of independence from Spain. Walker argues that the conspicuous lack of universal rights in the republican constitutions responds to the general distrust of the European elite of the indigenous population after the almost complete collapse of the colonial system as a result of Tupac Amaru's and Tupac Katari's uprisings.

How important are these types of (mis)perceptions in determining group composition and economic outcomes?[3]

We argue that the study of racial discrimination in contexts other than the U.S. is interesting, primarily because a multiracial setting, such as in Peru, makes it possible to study the meaning of race and how it interacts with discrimination. Anthropologists argue that race in Latin America is based on phenotypical (appearance) rather than genotypical (ancestry) characteristics.[4,5] Therefore, the measurement or determination of race is important. If race were a culturally-defined trait, it would not only be difficult to detect the presence of discrimination due to measurement problems, but it would make it difficult to design policies that successfully target disadvantaged populations that are not simultaneously susceptible to fraud. For instance, racial discrimination might manifest itself as a beauty premium, which is harder to detect and monitor.[6] In segregated societies in general, it is difficult to observe interactions among certain segments of the population at all. Even if interactions are observed, they will likely be limited and hierarchical as different populations are sorted into different occupations.[7] Considering such limitations, if perfect sorting into professions is observed, there is little hope in saying much about the extent of discrimination without resorting to strong exclusion restrictions.[8]

[3] There is little economic research aimed to detecting discrimination in Peru. Nopo, Saavedra and Torero (2004) study wage gaps between white and indigenous workers and find it to be around 12%. They notice also that these differences in earning were present only in the formal sector. Intriguingly, Moreno, Nopo, Saavedra and Torero (2004) do not find robust differences in the probability of being hired for job seekers of different ethnic backgrounds in a small audit study in Lima, Peru. While the authors consider that this apparent contradiction might be due to the sample size of the audit study, it is clear that much is still to be learned about the roots of differential life outcomes in Peru.

[4] Goldsmith, Hamilton, and Darity (2005) show evidence against the idea that race in America is a cultural trait by showing that light skinned blacks do not experience wage gaps as brown-skinned and dark blacks do. Gyimah-Brempong and Price (2005) show that skin tone also affect transition into crime and sentence durations. Darity, Dietrich, Hamilton (2005) argue against the idea that race in Latin America is phenotypical. They present evidence of strong preferences for whiteness among people of mixed blood.

[5] The fact that racial mixing and cultural adaptation are potential strategies to escape discrimination makes the issue of measuring race the most salient.

[6] There may be some truth to this, in that the Peruvian government passed a law in 1999 forbidding the requirement of "good presence" for clerical positions. This requirement could be interpreted as code for “not being indigenous.”

[7] See Moreno, Nopo, Saavedra and Torero (2004) for a discussion of gender and race occupational sorting into professions in Peru. Blau and Ferber (1992) show that occupational sorting is much more pronounced in Latin America than in other regions.

[8] Castillo and Petrie (2004) exploit the fact that a recent civil conflict in Peru pitched populations with different backgrounds to show that human right abuses were not random.

This research uses laboratory experiments to examine how and why people use observable characteristics and salient performance signals to sort into groups. People may focus on personal characteristics, such as sex, race, or beauty to choose group members because they lack better information on future performance (i.e. statistical discrimination). If personal characteristics and performance are correlated (or thought to be), it is difficult to determine whether personal characteristics have an impact on who is chosen as a group member. For example, if men are better performers and at the same time a preference for men as group members is observed, are men chosen as group members because of being better performers (statistical discrimination) or because people prefer to be in groups with men (taste-based discrimination)?

Our experimental design addresses the issues raised above.[9] First, our experiments test the hypothesis of the existence of taste-based discrimination by manipulating the information made available to subjects and by inducing behavior to break the correlation between performance and personal characteristics. We do this by making personal characteristics irrelevant or bad predictors of behavior and therefore irrelevant to payoffs. We argue that, regardless of the fact that a subject's expectations on the future behavior of others might be correct or not or might be observed by the researcher or not, taste-based discrimination should be *immune* to information on performance. That is, taste-based discrimination does not suggest that people disregard information on appearance once information on performance is revealed. Our experiments can test this. We consider this approach -- manipulating information and performance to test the nature of discrimination -- to be one of the strengths of our design since measuring expectations is not a trivial task (see Manski, 2004).

We use a repeated linear public goods game to explore these issues. Repeated public goods experiments represent a natural environment to study group formation because payments in the experiment are a function of both individual and group behavior and mimic many social situations. The more cooperative are other group members, the more money a person makes. In our experiment, subjects are asked to choose who they would like to have in their group in a surprise task before the last rounds of play. Treatments determine the type of information made

[9] The design was first presented in Castillo and Petrie (2006).

available to subjects. Subjects are shown either digital photographs of others in the experiment or information on past performance (or both).

Because performance and appearance may be correlated, it is important that our experimental design addresses this issue. If they are, simply providing information on appearance and performance will not be enough to identify which element affects sorting. To observe the importance of appearance on sorting we need a counterfactual situation where behavior contradicts held beliefs. So, one of our treatments breaks the correlation between performance and appearance, and this allows us to identify discrimination for other than statistical reasons. Our approach is novel in that it manipulates information or the equilibrium at the experimental level within the same game to identify sources of discrimination. Our environment is strategic and relevant to understanding how groups or neighborhoods form.

Second, the study develops a measure of race based on intensity of a genotype (Torero et.al 2004; and Nopo, Saavedra and Torero, 2005).[10] Separate questions were asked about how much White, Indigenous, Black, or Asian a person is. This task was performed by subjects not involved in the experiment, but recruited from the same population as the experimental subjects. In addition, another set of raters were recruited and trained to identify genotypical features from the pictures of subjects and ignore any feature related to looks or dress. As discussed later in the paper, the results are robust to either method used. Additionally, another set of subjects was recruited to rate experimental subjects according to physical attractiveness.

Finally, in an attempt to observe non-student populations that normally do not interact perform the same task, we recruited among the working population between 20 and 35 years of age in Lima, Peru. The sample, while small, is similar to the population at large. Moreover, by restricting ourselves to the working population we diminish the common criticism that student populations might be quite different than the general population (especially in developing countries where college education is uncommon) and might bias the results towards no discrimination.[11]

[10]Similar techniques were used before by Angel and Gronfein(1988) and Anderson, Silver and Abramson (1988).
[11] Harrison and List (2004) discuss these issues extensively.

There have been several studies trying to discern the nature of discrimination. Khan (1991) presents evidence of wage discrimination in basketball but not in baseball in the U.S. Audit studies suggest findings that are consistent with taste-based discrimination (Riach and Rich, 2002), but there are concerns about treatment effect biases (Heckman, 1998). Bertrand and Mullainathan (2004) find that those with black-sounding names tend to be discriminated against in a study using fake resumes. Knowles, Persico and Todd (2001) develop a test of taste-based discrimination in police car searches. They find evidence of statistical discrimination but not taste-based discrimination.[12] A more robust test of taste-based discrimination was suggested by Anwar and Fang (2006). They also find evidence of statistical but not taste-based discrimination. Levitt (2004) exploits the changes in incentives in the *Weakest Link* television show to test for alternative theories of discrimination. He does not find evidence of race or gender discrimination but of age discrimination. Finally, List (2006) finds evidence of age discrimination in choosing partners in the television show *Friend or Foe*.

The experimental literature has also addressed the issue of discrimination. Fershtman and Gneezy (2001) show evidence of statistical discrimination in Israel. They observed that people mistrusted men of Eastern origin, but otherwise did not make a difference when given the opportunity to make transfers to them. List (2004) also provides evidence of statistical discrimination in a sport cards market by collecting additional evidence with experiments. He finds that difference in bargaining behavior can be explained by difference in the distribution of reservation valuations and willingness to pay. Similar to Fershtman and Gneezy, he uses allocation exercises to test for taste-based discrimination and finds no evidence of it. Our study differs in that we test for discrimination by manipulating the equilibrium at the experimental level and exploit the racial diversity of Peru to investigate the measurement of race and its interactions with discrimination.

We find that the answer to the issue of racial discrimination is a complicated one. People do use others' personal characteristics to make economic decisions. However, attractiveness, rather than

[12] Taste-based discrimination refers to Gary Becker's (1975) taste-based theory where lenders relinquish profits in order to avoid interaction with specific groups.

race, is a much better predictor of unequal treatment. Our estimates of the effects of others' appearance on behavior are large and robust. A person considered unattractive is 12% less likely to be chosen as a member of the most-preferred group, and a person considered attractive is 29% more likely to make it to this group. We also find that race and attractiveness are strongly correlated. The probability of being considered unattractive given that a person is indigenous is 78%, but only 22% for a person classified as white. Once information on performance is provided, most evidence of discrimination is eliminated. However, our experiments suggest that the remaining differences are due to taste-based and not statistical discrimination.

Our study shows the advantage of experimental methods in tackling difficult identification issues. It also shows the importance of measurement of personal characteristics and sampling in the study of race and beauty in experiments.

The Sample

The experiments were conducted in urban metropolitan Lima in Peru. We chose this site because we want a broadly representative, non-student sample of the population that is familiar with computers and the internet. This is essential since our experiments are internet based. Lima is replete with internet cafes, and there is a high proportion of the non-student population with expertise using the internet. By drawing upon this broader population, we are able to look more accurately at the extent of discrimination.

Sample Selection Procedure

Our sampling strategy is consistent with the goals of the experiment. First, we want to create an environment in which people of various social distances who might not normally interact with one another can. Second, at the same time, we want to have a sample of subjects which is representative of the young working population in metropolitan Lima. With these objectives in mind we implement the following strategy.

First, we define an eligibility criterion for the subjects. To be a subject she/he must be between 20 and 35 years of age, live in Metropolitan Lima, have labor market experience, be currently working, know how to use the Internet, and have an e-mail account. In addition we sought to keep a gender and income balance so that subjects would be distributed homogenously across gender and income levels.

Second, based on these criteria we contracted two companies specialized in surveys and recruiting. The main asset of these companies is that they maintain representative databases for metropolitan Lima. This helped us to secure a diverse population in the experiments. In general, this mechanism ensures that the opportunity to participate in the experiment is distributed equally across the population. From these databases we sampled all the potential subjects that comply with all of our criteria. From the resulting sub-sample we performed a random lottery and selected the individuals to be part of the experiment.

Finally, and with the interest of over sampling clusters of owners of small, medium and micro-enterprises, we sent a team of recruiters to Gamarra (an industrial area in metropolitan Lima) to randomly select subjects.[13] The advantage of Gamarra was two fold. First, we had a pre-census of all the establishments in Gamarra and this allowed us to randomly select buildings from which to invite subjects. Second, it is one of the biggest small- to medium-sized enterprise clusters in metropolitan Lima and represents a rich mix of population in terms of place of origin and socio-economic background.

The protocol used for the experiments was simple enough to include large segments of the population. The interface was graphical and required simply that the subjects know how to use a computer mouse. It is important to note, however, that because our experiments rely on internet protocols and the knowledge of using a computer, we likely excluded some segments of the population that might suffer more marked patterns of discrimination. Therefore our results give a lower bound estimate to the extent of discrimination.

[13] First, based on a pre-census, we identify the blocks with the highest concentration of commercial buildings. Second, we visit 4 commercial buildings per block randomly selecting one and from it one each 3 buildings to the left or the right. Third, within the buildings we select the establishments by performing a random lottery from the total number of establishments in each odd level of the building. Fourth, we invite one person with the establishment to participate in the experiment.

Comparison of Sample to Metropolitan Lima Population

Figure 1 gives an approximation of the distribution of the selected subjects across metropolitan Lima. According to the population census of 1993, our sample essentially covers most of the districts in Metropolitan Lima and is highly correlated with the distribution of the population with complete or incomplete higher education.

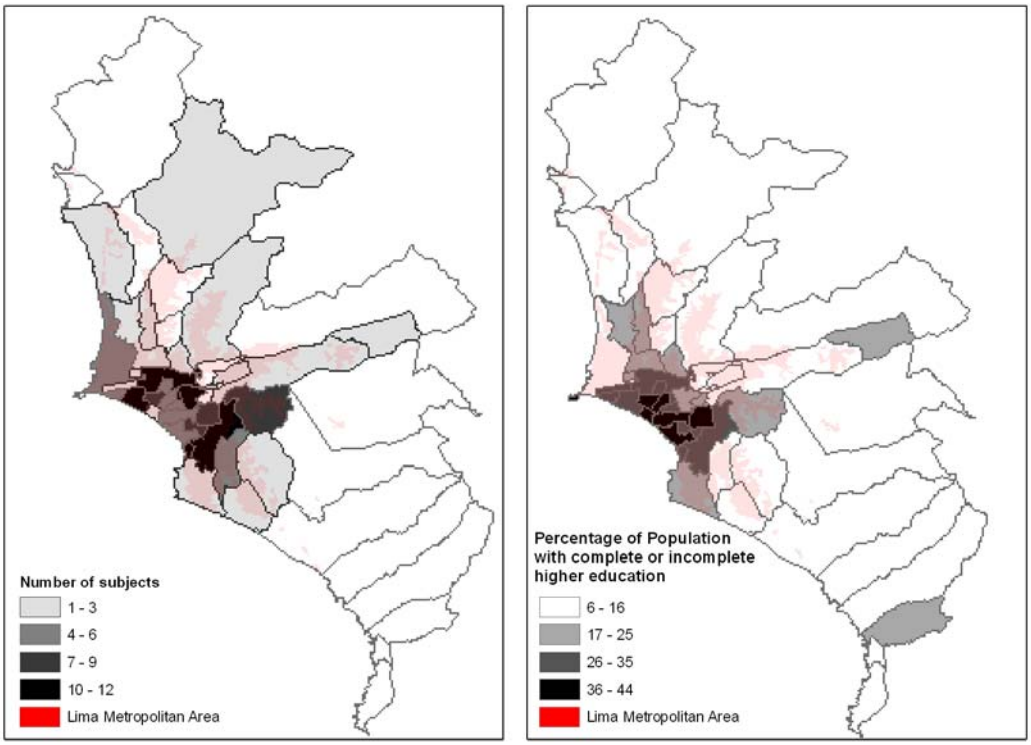

Figure 1: Distribution of the sample in comparison with the population with complete or incomplete higher education in the population census of 1993

To investigate the comparability of our sample to the population in other dimensions, we compare out experimental subjects to a sub-sample from the *Encuesta Nacional de Hogares* (ENAHO) 2004. The sub-sample complies with the eligibility criteria for all of our subjects. The advantage of using the ENAHO as a comparison group is that it is representative of Metropolitan Lima and therefore could help us identify any selection bias in our sample. Table 1 presents the results of comparing our experimental subjects and ENAHO 2004 for a subset of variables

common to our post-experiment questionnaire and to the ENAHO. Both groups have a similar distribution among almost all the variables, but our experimental subjects are slightly more educated. This is most likely a reflection of the requirement in our experiment that subjects know how to use the internet. However, the age, gender, monthly income, average education, and language distribution are very similar. This comparison is important in that it gives us confidence that the subjects in our experiment are a good representation of the larger population in metropolitan Lima.

Table 1: Comparing our Experimental Subjects with ENAHO 2004

	Experimental Subjects	ENAHO 04 [1/.]
Age	26.30	27.95
Male	61.2%	54.6%
Level of education [2/.]		
Incomplete non-university tertiary	23.1%	24.6%
Complete non-university tertiary	15.6%	36.2%
Incomplete university tertiary	31.9%	20.0%
Complete university tertiary	29.4%	19.2%
Years of education [2/.]	15.10	14.02
Native language [3/.]		
Quechua	2.5%	1.5%
Spanish	97.5%	98.5%
Monthly income [4/. 5/.]	1087.3	1099.2

1/. Estimations for ENAHO 2004 have been limited to observations with the following characteristics: (a) residence in Metropolitan Lima; (b) 20-35 years old; (c) working at the time of the survey; (d) with more than complete secondary education.

2/. Some respondents in the sample reported less than complete secondary education and have been excluded from schooling-related estimations.

3/. This information is only available for the III and IV Quarter Rounds of ENAHO 2004.

4/. Some respondents in the sample reported they were not working at the time of the survey and have been excluded from labor-related estimations.

5/. In constant soles of May 2006.

Experimental Design

We use a linear public goods game to explore discrimination in group formation. This design was first developed and used by Castillo and Petrie (2006). Each subject is given a 25 token endowment and must decide how to divide the endowment between a private investment and a public investment. Each token placed in the private investment yields a return of 4 centimos to the subject.[14] Each token placed in the public investment yields a return of α_i to the subject and every other member of the group. The return to the public investment, α_i, is 2 centimos in three of the four treatments. There are 20 subjects in each experimental session. Subjects are randomly assigned to a five-person group and play 10 rounds with that same group. At the end of each round, subjects learn their payoff, π_i, and the total number of tokens contributed to the public investment by the group, G. In total, subjects play three 10-round sequences, and each 10-round sequence is with the same group. At the end of the first 10-round sequence, subjects are again randomly assigned to a new five-person group, and at the end of the second 10-round sequence, subjects are asked to choose their group for the final 10 investment decisions. Subjects do not know they will be asked to choose their group before this point in the experiment.

In order to create an incentive for people to reveal who they would prefer to be in their group, we create the following procedure. Subjects rank all the other 19 subjects in the session from 1 (most preferred) to 19 (least preferred). We provide subjects with some information on the other subjects in the room to use for ranking. The information is either the average amount contributed to the public investment during the second 10-round sequence, the subject's photo, or both. Subjects use that information to create a list from most preferred to least preferred. Digital photographs of subjects are taken at the beginning of the experiment, and photographs are head shots, similar to a passport or identification photo.

Once all subjects submit their lists, groups are formed in four steps. First, one person is chosen at random. A group is formed that includes the randomly chosen person and the top four people on his list. Second, one person from the remaining 15 people who have not been assigned to a group is randomly chosen. A group is formed with that person and the first four people on that person's

[14] There are 100 centimos in 1 sole (the Peruvian currency). At the time of the study, US$1 = 3.2 soles.

list from the remaining people who have not been previously assigned to a group. Third, one person from the remaining 10 people who have not been previously assigned to a group is randomly chosen. The first four people on that persons list among the remaining people are put in a group with that person. Fourth, anyone not already assigned to a group is put in a group together. Once groups are formed by the procedure described above, subjects then see a screen with the information corresponding to the subjects in their new group. Subjects play the last 10 rounds with that group.

This mechanism is similar to the one suggested in Bogomolnaia and Jackson (2002). The mechanism is incentive compatible if preferences over groups are additive in the preferences over its members. Additivity in this context means that if Pablo prefers Maria's company to Gabriela's company, then Pablo always prefers a group than exchanges Gabriela for Maria, regardless of who the other members of the group are. Under these conditions, revealing the ordering of others is a weakly dominant strategy for Pablo. If Pablo is not chosen, he is indifferent in the ranking he reveals, but if he is chosen, he is better off by revealing his true rankings. Since preferences over others' company is additive, it does not matter whether he is chosen first or last.

Some may argue that additivity of preferences over others' company may be a strong assumption. Some combinations of people might be less successful than others. For instance, women might be very cooperative with other women but not so with men. Therefore, a woman might be chosen to be part of a group when other women are available, but not when mostly men are available.

There is another mechanism that is incentive compatible, regardless of preferences over groups. If people are able to rank all possible groups that one could be paired with, we would not need to be concerned with the additivity assumption. Unfortunately, this option would be impractical since the number of groups to be ranked would be exceedingly large.[15] For this reason, we opted for the mechanism described above because it is easy to explain to subjects and can be implemented quickly once subjects have submitted their lists of rankings.

[15] With 20 subjects, each subject would need to rank 3,876 groups.

There are four experimental treatments: Contribution Only, Photo Only, Contribution and Photo, and Two Types. Treatments differ in the α_i assigned to each person and the information that is shown to subjects when they are asked to rank the other subjects.

In the Contribution Only, Photo Only and Contribution and Photo treatments, all subjects are assigned an $\alpha_i = 2$ centimos, so the price of contributing to the public good is 2. It is in the group's interest for everyone to contribute their full endowment to the public investment, but each individual in the group maximizes his own payoffs by putting all his tokens in the individual investment. In the Contribution Only treatment, when subjects are asked to rank others, they see the average amount contributed to the public good in the second 10-round sequence by all other subjects in the room. In the Photo Only treatment, subjects see the photos of all other subjects. And, in the Contribution and Photo treatment, subjects see the photo and the average amount contributed to the public good in the second 10-round sequence. The average is listed below each subject's photo.

In the Two Types treatment, $\alpha_i \in \{0.5, 5.0\}$ centimos. Half of the subjects are randomly assigned a value of 0.5 and half are randomly assigned a value of 5.0. Subjects keep the same value for all 30 rounds of play. A subject with an $\alpha_i = 5.0$ has a price of contributing to the public good of 0.8 and should invest his entire endowment in the public good, whether he is selfish or altruistic. A subject with an $\alpha_i = 0.5$ has a price of contributing to the public good of 8, so investing in the public good is very expensive. We would expect subjects assigned the low α_i to invest little to nothing in the public good. Because subjects are randomly assigned a type, performance and appearance will not be correlated. The Two Types treatment is key to our ability to identify whether appearance or performance affects sorting.

Each treatment was run twice, and each experimental session had 20 subjects. An experimental session lasted at least two hours. In total, 160 subjects participated in the four treatments. Each session ended with an extensive post-experiment questionnaire. The experiments were conducted on computers in two computer labs at the Pacific University in Lima, Peru. Two treatments were run at the same time, so subjects were randomly assigned to treatments. Since most subjects worked full time, the experiments were conducted on weekend afternoons.

In the Contribution Only, Photo Only, and Contribution and Photo treatments, average payoffs are $19.65 (standard deviation $1.36). In the Two Types treatment, average payoffs are $33.75 (standard deviation $6.87).

Race and Beauty Classifications

We are interested in knowing if people sort into groups based on physical characteristics. While a person's sex is easy to determine, a person's race or beauty is not. We want to develop an independent measure of the race and beauty of a person that reflects the general perception of that person. Therefore, we use raters, people who did not participate in the public goods experiment but who are drawn from the same cohort as the subjects in the experiment, to rate the photos of the subjects in terms of race and beauty.[16] A rater only rated the photo in terms of one characteristic, race or beauty, not both.[17]

For race ratings, because the most popular self-classification of race in Peru is *mestizo* (mixed race), it is important for us to have a measure of race that can adequately capture this mixing. For this reason, we use the race classification method developed by Torero et al. (2004) and Nopo et al. (2005). Instead of classifying a subject along one dimension of "white" or "mestizo," we evaluate subjects in their racial intensity in four categories: white, indigenous, black and asian. These are groups that people readily recognize as distinct racial groups. This gives a more nuanced measure of race and more accurately captures the racial mixing in Peru.

To obtain these ratings, we had twenty people (10 women and 10 men), not involved in the public goods experiment, rate each subject along each of these four dimensions. Each dimension was rated from zero to ten, with zero being complete absence of the dimension and 10 being the most intense. Raters were instructed to choose whichever number between zero and 10 best

[16] This technique has been used in other experimental research on beauty, including Andreoni and Petrie (2006), Eckel and Wilson (2003), Mobius and Rosenblat (2006). Hamermesh and Biddle (1995) had interviews rate the interviewees in terms of beauty.

[17] Half of the raters were men, with an average age of 27.4 years. For education, 21.7% had incomplete non-university tertiary, 16.7% had complete non-university tertiary, 25.0% had incomplete university, and 21.7% had complete university education.

described the person for each of the four racial dimensions. The four numbers did not need to add up to 10. The raters were also told that if they thought that a person belonged to only one racial group, then that person should be given a 10 for that racial dimension and a zero for all other dimensions. Raters were shown the photos one by one on a computer screen and chose the intensity of each dimension by clicking on a button. Raters could easily move back and forth between the photos to check or change their answers. Ratings took about one hour, and each rater was paid $9.38 (30 soles) for their time.

For the beauty rating, we followed the same procedure as with the race ratings. The only difference is that the ten men and ten women were asked to rate the physical attractiveness of each person in the photo on a scale of one to nine, with one being very unattractive and nine being very attractive. There was a high degree of agreement among raters in terms of attractiveness. Pairwise correlations among raters ranged from 0.13-0.75, with an average of 0.50.[18]

In terms of agreement among raters on race, there was also a high degree of agreement in terms of each racial dimension. Along the white dimension, pairwise correlations among raters ranged from 0.31-0.76, with an average of 0.57. For the indigenous dimension, correlations ranged from 0.02-0.64, with an average of 0.41. For the black dimension, correlations ranged from 0.19-0.82, with an average of 0.50, and for the asian dimension, correlations ranged from -0.02-0.81, with an average of 0.37.[19]

Note that we also had trained raters, in addition to our cohort raters, rate the photos in terms of racial intensity. These raters were trained to minimize variance in racial perceptions. There was a large amount of agreement between the trained raters and cohort rates. For example, along the indigenous dimension, pairwise correlations ranged from 0.20-0.79, with an average of 0.55, and along the white dimension, pairwise correlations ranged from 0.27-0.78, with an average of 0.57. This indicates to us that race can be measured and defined. It is clearly observable, and people

[18] The Cronbach alpha for interrater reliability is 0.94.
[19] The Cronbach alpha for interrater reliability 0.9565 for the white dimension, 0.9285 for the indigenous dimension, 0.9451 for the black dimension, 0.9113 for the asian dimension.

can clearly use it to discriminate. This gives us confidence that the race variables we are using are actually picking up the effects of race.

While there are some subjects that display intensities in the dimensions of black and asian, the majority of subjects display the greatest intensities in the dimensions of white and indigenous. This is in line with the general population in Peru, where blacks make up 2% of the population and Asians make up 3% of the population. Average intensity is 2.83 for white, 3.91 for indigenous, 1.89 for black, and 1.31 for asian. Because the majority of our subjects were primarily a mix of white and indigenous, we concentrate on these two dimensions in our analysis. None of our analysis changes if we add asian and black intensities.

While the rating scale for race ranged from zero to ten and for beauty from one to nine, some raters did not use the full range of the scale. For example, for race, some used intensities up to 10 and some only up to 6. To be able to make comparisons across raters, we standardize each rater's rating by her own mean and standard deviation. This permits us to take an average across all twenty raters' standardized ratings for race and for beauty to get the final average ratings we use to analyze the data.

In lieu of using average intensities of race or beauty, we create dummy variables for our analysis. A person is classified as White if the average standardized rating for that person in the white racial dimension falls in the top tercile of the distribution *and* the rating in the indigenous dimension falls in the bottom tercile of the distribution. A person is classified as Indigenous if he falls in the upper tercile of the indigenous distribution *and* in the lower tercile of the white distribution. Given this definition, 25.6% of subjects are classified as White and 22.5% are classified as Indigenous.

For beauty, women are rated as more attractive than men. The average standardized attractiveness measure for women is 0.35 and -0.22 for men. Therefore, we classify subjects as attractive or unattractive, conditional on their sex. So, a man is classified as attractive if his average standardized attractiveness rating falls in the upper tercile of the distribution of attractive ratings for men. And, a man is classified as unattractive if his rating falls in the lower tercile of

the distribution of ratings for men. The same procedure is used for a woman, conditional on how her rating falls in the distribution of ratings for women.

Results and Discussion

What Did People Do in the Experiment?

Figure 2 shows the average behavior by round across treatments for sequence two. As it is commonly observed in public goods experiments (see Kagel and Roth, 1995), contributions tend to decline over time. Contributions in all treatments, except Two Types, start around 30% and decline to around 15% in the last round. A similar pattern is also observed in the first sequence of the experiment, although not shown in the paper. Low types in the Two Types treatment contribute about 30% and decline slightly in the last few rounds, and high types contribute roughly the same in early and late rounds.

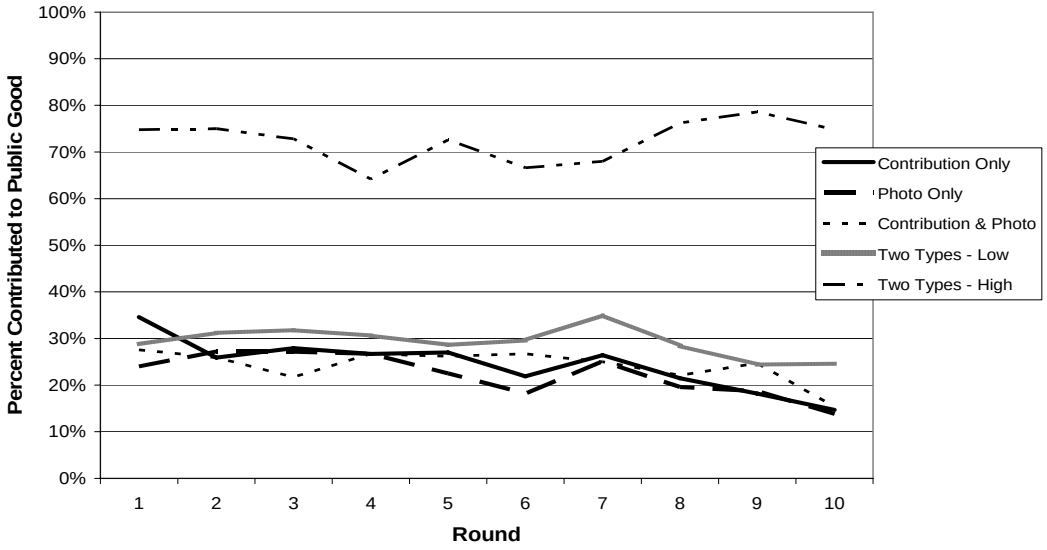

Figure 2: Contributions to the Public Good, Second Sequence

The figure also shows that the incentives of the Two Types treatment successfully induce a separation in behavior between high and low types. High types contribute about two and a half times as much to the public good than low types. They contribute about 79% over all rounds, and

low types contribute about 30%. This provides the split in behavior, uncorrelated with appearance, that we need to test discrimination.

A basic premise in theories of statistical discrimination is that, in the absence of better information, ethnic or cultural background can be used as a proxy for behavior. For instance, migrants might experience rough market conditions, making them behave (or thought to behave) more selfishly. Or, more affluent subjects can afford to be more altruistic or take more risks. Table 2 shows a series of random-effects regressions aimed at determining if different people do behave differently. All regressions include group-level fixed effects in order to control for the fact that different levels of contributions might be observed due to social interactions within a particular group. The regressions also include random effects at the individual level to control for the fact that the same person's decisions are correlated.

Included in the regressions are variables for sex, age, education, race, beauty and assigned type in the Two Types treatment. The variables for race and beauty are dummy variables constructed from the average standardized continuous measures as described in the previous section.

Table 2: Percent of Endowment Contributed to the Public Good
Group-Level Fixed Effects and Individual Random Effects Regression (Sequence 2)

Variables	Contribution Only, Photo Only, & Contribution and Photo	Two Types	All Treatments
Male	4.52 (0.11)	10.66 (0.18)	6.00 (0.03)
Age (years)	0.12 (0.75)	-0.67 (0.45)	-0.11 (0.75)
Education (years)	0.40 (0.66)	-1.78 (0.39)	-0.40 (0.64)
White	0.09 (0.98)	-6.08 (0.61)	-0.64 (0.87)
Indigenous	-0.44 (0.92)	-3.50 (0.77)	-1.07 (0.80)
Attractive	-1.14 (0.77)	8.20 (0.53)	0.87 (0.82)
Unattractive	-1.56 (0.71)	6.86 (0.51)	1.06 (0.79)
Low Type			-7.90 (0.45)
High Type		46.19 (0.00)	35.46 (0.00)
Round	-1.19 (0.00)	-0.12 (0.73)	-0.92 (0.00)
Constant	25.49 (0.15)	60.73 (0.13)	47.87 (0.00)
Individual Random Effects	*Yes*	*Yes*	*Yes*
Group Fixed Effects	*Yes*	*Yes*	*Yes*
Within-R2	0.04	0.00	0.02
N	1200	400	1600
p-values in parentheses			

The regressions in Table 2 show that behavior is essentially not correlated with personal characteristics. On average, contributions decrease by 10% from round 1 to round 10, and there is an effect of men giving 6% more in the pooled regression. Comparing the column showing results from the combination of the Contribution Only, Photo Only, and Contribution and Photo treatments and the column showing the results for all treatments, we see that high types in the Two Types treatment contribute 35.5 percentage points more than Low Types.

Table 2 shows that personal characteristics are of little help in predicting others' behavior. While men contribute slightly more in the pooled regression, this does not hold in all specifications. It is important to note at this point that Table 2 shows that race and beauty are not correlated with behavior at all. The weakness of personal characteristics as an explanation of behavior will be useful in interpreting the results in the following section.

How Were People Ranked?

We have seen that personal characteristics explain little, if any, of behavior. But, are personal characteristics used when choosing groups? Table 3 reports individual ranker's fixed effects regressions of ranking on age, sex, race, beauty and expected rank for each treatment separately.[20] The dependent variable is the rank that a person is given. A person with a rank of 1 is ranked highest and a person with a rank of 19 is rank lowest. This means that if a coefficient is positive then the variable associated with it tends to lower one's rank. If a coefficient is negative the presence of the covariate tends to improve one's rank.

The race and beauty dummy variables are the same as were used in the regression in Table 2. There is, however, one covariate that requires extra explanation. Expected Rank is a variable indicating the rank that a person should have if only contributions to the public good are used to rank others. This means that if a person had the highest average contribution in sequence two in that session, then the expected rank would be one. The lowest contributor has an expected rank

[20] The results in Table 3 are robust to alternative estimations: OLS with clustered errors instead of fixed effects and random-effects Tobit. The results are also similar if we use racial intensities of trained raters, instead of cohort raters.

of 19, and any ties are assigned the average rank. The expected coefficient on this variable should be 1 if information on others' behavior is the only relevant information in creating ranks. All regressions include fixed effects of the person doing the ranking since each individual ranked 19 people.

Table 3: Fixed Effects Regression on Individual Ranking (highest = 1, lowest = 19)						
Variables	**Photo Only**		**Contribution & Photo**		**Two Types**	
Age (years)	-0.00	-0.01	0.04	0.03	-0.02	-0.04
	(0.96)	(0.81)	(0.21)	(0.30)	(0.52)	(0.31)
Male	1.98	1.61	0.06	0.08	-0.03	0.23
	(0.00)	(0.00)	(0.81)	(0.78)	(0.92)	(0.48)
White	-1.85	-1.35	-0.20	-0.24	-0.03	0.54
	(0.00)	(0.02)	(0.46)	(0.46)	(0.94)	(0.29)
Indigenous	1.20	-0.84	0.20	-0.17	0.17	-0.40
	(0.02)	(0.20)	(0.49)	(0.63)	(0.66)	(0.37)
Attractive		-0.89		0.04		-0.85
		(0.10)		(0.92)		(0.08)
Unattractive		2.35		0.63		0.63
		(0.00)		(0.06)		(0.12)
Expected Rank			0.83	0.82	0.66	0.67
			(0.00)	(0.00)	(0.00)	(0.00)
Constant	8.83	9.12	0.66	0.78	4.01	4.27
	(0.00)	(0.00)	(0.46)	(0.41)	(0.00)	(0.00)
Individual Fixed Effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Within-R2	0.06	0.09	0.71	0.71	0.44	0.44
N	760	760	760	760	760	760
p-values in parentheses						
Note that the above results are robust to specifications including other variables such as education, height, high school attended, mother tongue and religion.						

People seem to understand that having high contributors in the group is the best strategy. For instance, expected rank alone explains 67% of the variance of ranks in Contribution Only (not shown in Table 3). Expected Rank remains a strong predictor of rank, and explains a large part of the variation, in all treatments where information on previous contribution is provided.

Interestingly, despite the fact that personal characteristics have no bearing on the contribution choices of people in the experiment, they tend to predict the way people are ranked in the Photo Only treatment. Men are ranked on average 1.6 to 2.0 ranks lower than women. Without

controlling for beauty, white subjects are ranked 1.8 ranks higher and indigenous subjects are ranked 1.2 ranks lower. When beauty is added, Whites are still ranked higher, but indigenous subjects are not ranked lower. It is unattractive subjects that are ranked 2.4 ranks lower. Being attractive does not seem to help in the rankings, but being unattractive really hurts. Beauty also helps explain another 3% of the variation.

Who is doing the discriminating in the Photo Only? Both men and women rank men and unattractive people lower, but it is women who rank Whites higher. Both Whites and Indigenous rank unattractive people lower. White subjects rank Whites higher, and indigenous subjects rank men lower.

It is important to note that race and beauty are highly correlated. Seventy-eight percent of indigenous subjects are classified as unattractive, and 78% of white subjects are classified as attractive. This is extremely telling, since this is a result of the intersection of two separate and independent sets of raters, one for race and one for beauty. In essence, it is not that one rater perceives the subject to be both indigenous and unattractive, but one rater perceives the subject to be indigenous and another rater perceives him to be unattractive. Combining the two ratings, we see that the majority of indigenous subjects are also classified as unattractive. For attractive subjects, it is mainly their whiteness that boosts their rankings, but for indigenous subjects, it seems that it is their lack of beauty that affects their rankings.

While race, beauty and sex affect rankings in Photo Only, rankings in treatments where information on past performance is available are affected only by performance and beauty. In Contribution and Photo and Two Types, a large percent of the variation, between 36-61%, can be explained by expected rank. Beauty is significant, in that unattractive people are ranked lower in Contribution and Photo and attractive people are ranked higher in Two Types. Because the beauty ratings are gender specific, this means that both attractive men and attractive women are ranked higher in Two Types.

It is important to recall the purpose of the Two Types treatment. If contribution behavior and personal characteristics are correlated, then the Contribution and Photo treatment will not allow

us to cleanly measure which variables, contribution or characteristics, affect rankings. The Two Types treatment allows us to do so by breaking the correlation between performance and characteristics. The results from the Two Types treatment give us confidence that both beauty and performance affects rankings, even though, by the experimental design of Two Types, beauty is orthogonal to performance. Indeed, the results from Tables 2 and 3 together suggest that there is some stereotyping or taste-based discrimination.

Is it stereotyping or taste-based discrimination? According to Phelps (1972), statistical discrimination can be thought of as an error-in-variables problem. Subjects might judge similar evidence on performance differently if either behavior of some groups is more variable or subjects have less informative priors on others behavior as social distance increase. Regarding the latter explanation, there is no evidence that different groups of people doing the ranking are any more likely to rank attractive people higher, so it does not appear to be an issue of social distance. Regarding the former explanation, the performance of attractive subjects is slightly less variable. Only men's behavior is more variable.[21] The variability of attractive versus unattractive subjects was common knowledge since all subjects saw the contributions of all subjects and their pictures.

Statistical discrimination theories would suggest that the attenuation bias on the parameter associated with expected rank will also be associated with appearance. Indeed, regressions, shown in the Appendix, with interaction terms for attractiveness and expected rank show that the impact of expected rank is smaller for attractive subjects. This means that people are more likely to disregard the same performance from an attractive person than from an unattractive person. This seems to contradict statistical discrimination. Because attractive subjects are *less* variable in their behavior, we would expect people to pay more attention to their behavior when ranking rather than disregarding it.

Risk aversion would predict that attractive subjects would be ranked higher if their behavior is less variable. However, risk aversion would also predict that men, whose behavior is more variable, be ranked lower. They are not. This leaves us with the conclusion that the differential

[21] This was also found by Andreoni and Vesterlund (2001) and Andreoni and Petrie (2006).

ranking for attractive people in the Two Types treatment is more likely due to taste-based discrimination, rather than statistical discrimination.

It might be that the relationship between personal characteristics and ranking is non-linear across the full list of ranking. Subjects might pay more attention to the top and the bottom of the list. To investigate this, we run a logit model, with individual fixed effects, to see if the probability of making it to the top 4 or the bottom 4 on the ranking list is affected by personal characteristics and the expected probability of making it to the top or bottom group. The latter variable is constructed as follows. If the subject's expected rank was strictly less than 5, then the subject was assigned the value of one for the expected probability of making it to the top group. If the subject's expected rank was strictly greater than 15, then the subject was assigned the value of one for the expected probability of making it to the bottom group. Tables 4 and 5 show these results.

Race is not a factor in making it to the top 4, but beauty is. In Photo Only, an attractive person is more likely to be in the top 4, and an unattractive person is not. Men are also less likely to be in the top 4. When past performance is available, people who are higher contributors are more likely to be ranked in the top 4. In Two Types, unattractive people are less likely to be in the top 4.

Table 4: Probability of Making it to the Top 4, Logit Regression

Variables	Photo Only	Contribution & Photo	Two Types
Age (years)	-0.00 (0.82)	-0.03 (0.50)	0.04 (0.13)
Male	-0.54 (0.00)	-0.18 (0.63)	-0.53 (0.10)
White	0.35 (0.13)	-0.21 (0.60)	0.03 (0.94)
Indigenous	0.37 (0.28)	-0.47 (0.32)	-0.13 (0.66)
Attractive	0.38 (0.09)	0.49 (0.28)	-0.20 (0.56)
Unattractive	-0.94 (0.00)	-0.23 (0.59)	-0.79 (0.01)
Expected to be in Group (based on Expected Rank)		4.04 (0.00)	3.32 (0.00)
Individual Fixed Effects	*Yes*	*Yes*	*Yes*
Log-Likelihood	-308.30	-133.13	-217.94
N	760	760	760
p-values in parentheses			

Table 5: Probability of Making it to the Bottom 4, Logit Regression			
Variables	Photo Only	Contribution & Photo	Two Types
Age (years)	-0.02 (0.35)	0.06 (0.18)	-0.03 (0.30)
Male	0.40 (0.05)	0.31 (0.53)	-0.20 (0.37)
White	0.03 (0.92)	0.24 (0.70)	0.05 (0.89)
Indigenous	-0.33 (0.22)	-0.39 (0.48)	0.02 (0.94)
Attractive	-0.24 (0.39)	0.04 (0.95)	-0.28 (0.42)
Unattractive	0.99 (0.00)	0.85 (0.13)	0.40 (0.16)
Expected to be in Group (based on Expected Rank)		4.71 (0.00)	2.75 (0.00)
Individual Fixed Effects	*Yes*	*Yes*	*Yes*
Log-Likelihood	-312.74	-95.54	-224.04
N	760	760	760
p-values in parentheses			

For the bottom 4, beauty is also a significant predictor. An unattractive person is more likely to be in the bottom 4 in Photo Only, as are men. The only significant variable in Contribution and Photo and Two Types is the person's past contribution.

The results for the Two Types treatment are interesting because subjects are induced to behave quite differently regardless of their looks or background. Despite this, unattractive people are less likely to be named among the top 4. Given the correlation between looking indigenous and being unattractive, this may be another way of saying that looking indigenous decreases the likelihood of being in the top group.

Conclusions & Policy Implications

We present a series of experiments aimed at determining the nature and extent of discrimination in urban Lima, Peru. Subjects play a linear public goods game and, in a surprise, are allowed to sort into groups for the last rounds of play. Our experiments systematically manipulate the information available about others when sorting into groups. This allows us to examine what is more relevant to group formation, information on past performance or physical characteristics. Our design is unique in that we can clearly identify the effect of personal characteristics from

performance and do not have to rely on measuring expectations. We recruited a diverse sample of individuals currently working in the labor market to participate in the experiments.

Our experiments show that subject behavior is not correlated with personal characteristics, be it race, sex or beauty. However, people do use the personal characteristics of others when given the opportunity to choose partners. Our results are consistent with evidence of stereotyping or taste-based discrimination. Interestingly, our experiments also show a great reduction in discrimination or stereotyping once information on others' behavior is provided. This is good news, since it shows that discrimination can be greatly diminished if appropriate measures of performance are made available.

Subjects tend to prefer groups of women and white-looking people and dislike groups with unattractive people. While evidence of discrimination is almost completely eliminated by revealing information on others' behavior, there is still evidence that beauty and race are important factors even when information is revealed and performance is orthogonal to appearance. Intriguingly, while women and white-looking people are preferred in the absence of information, they are no more likely to make it to the top ranks when information is revealed. The effect of beauty, however, seems to be constant, even when subjects are given incentives that make their behavior and personal characteristics uncorrelated. Note that being unattractive and looking indigenous is highly correlated, so some effects of beauty may also be picking up effects of race.

Without information on others' behavior, not everyone uses others' characteristics in ranking in the same way. This suggests some form of stereotyping. While there is agreement across sex and race that women are more desirable partners and unattractive people are less desirable partners, the effect of race on rankings is basically explained by the behavior of women and white participants. Our results show that the remaining evidence of discrimination in the treatments where information on performance is available is hard to reconcile with theories of statistical discrimination.

Our research shows that understanding racial discrimination requires not only distinguishing its nature but also overcoming the problems caused by measuring race and sorting. Indeed, we show that racial discrimination might be masked as beauty premiums. The fact that people sort into professions requires that measurement of unequal treatment be done in tasks that are comparable and subjects that represent the population at large. Experimental methods can be used to tackle these difficult identification problems. Our design keeps the task constant, measures personal characteristics, and creates the necessary counterfactuals to identify the nature and extent of discrimination. We show that race, as well as beauty, is a discernible characteristic, and we find a large degree of agreements among people.

There are some important policy implications. People seem to have preconceptions of the behavior of others that create a barrier to access. That is, if people are excluded based on their appearance, those being excluded are denied the opportunity of showing what they are capable of doing. Given that once information is revealed most discrimination goes away, it seems that it would be recommendable to create opportunities for people to interact and to have independent measures of performance. For instance, professional accreditation might lower barriers to entry to those otherwise disadvantaged by eliminating the stereotypes associated with lower tier institutions or schools.

References

Anderson, B.A., Silver, B.D. & Abramson, P.R, "The effects of the race of the interviewer on race-related attitudes of black respondents in SRC/CPS national election studies. *Public Opinion Quarterly*, 52, 289-324, 1988.

Andreoni, J. and R. Petrie, "Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments," Working Paper, Georgia State University, 2006.

Andreoni, J., and L. Vesterlund, "Which Is the Fair Sex? Gender Differences in Altruism," *Quarterly Journal of Economics*, 116(1), 293-312, 2001.

Angel, R. and W. Gronfein, "The Use of Subjective Information in Statistical Models, *American Sociological Review*, 53(3), 464-473, 1988.

Anwar, S. and H. Fang, "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Review*, 96, 127-51, 2006.

Becker, G., *The Economics of Discrimination*, 2nd ed., Chicago, University of Chicago Press, 1975.

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

Blau, F. and M. Ferber, *The Economics of Women, Men, and Work*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1992.

Bogomolnaia, A. and M. Jackson, "The Stability of Hedonic Coalition Structures," *Games and Economic Behavior*, 38(2), 201-30, 2002.

Castillo, M. and R. Petrie, “Discrimination in the Warplace: Evidence from a Civil War in Peru,” Working Paper, Georgia Institute of Technology, 2004.

Castillo, M. and R. Petrie, “Discrimination in the Lab: Experiments Exploring the Impact of Performance and Appearance on Sorting,” Working Paper, Georgia State University, 2006.

Darity, W., J. Dietrich and D. Hamilton, “Bleach in the Rainbow: Latin Ethnicity and Preference for Whiteness” *Transforming Anthropology*, 13(2), 103-10, 2005.

Eckel, C. and R. Wilson, “Conditional Trust: Sex, Race and Facial Expressions in a Trust Game,” Working Paper, Rice University, 2003.

Escobal, J., J. Saavedra, M. Torero, “Los Activos de los Pobres en el Peru,” Group of Analysis for Development, working paper no. 26, 1998.

Fershtman, C. and U. Gneezy, "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, 116(1), 351-77, 2001.

Goldsmith, A., D. Hamilton and W. Darity, “Shades of Discrimination: Skin Tone and Wages” *American Economic Review, Papers and Proceedings*, May 2006.

Gyimah-Brempong, K. and G. Price, “Crime and Punishment: And Skin Hue Too?,” *American Economic Review Papers and Proceedings*, 96 (2), 246-250, May 2006.

Hammermesh, D. and J. Biddle, “Beauty and the Labor Market,” *American Economic Review*, 84(5), 1174-94, 1994.

Harrison, G. and J. List. "Field Experiments," *Journal of Economic Literature*, XLII (December): 1013-1059, 2004.

Heckman, J., "Detecting Discrimination," *Journal of Economic Perspectives*, 12, 101-16, 1998.

Kagel, J. and A. Roth, *Handbook of Experimental Economics*, 1995.

Kahn, L. "Discrimination in Professional Sports: A Survey of the Literature," *Industrial and Labor Relations Review*, 44, 395-418, 1991.

Knowles, J., Persico, N., and P. Todd. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy*, 109, 203--29, 2001.

Levitt, S., "Testing Theories of Discrimination: Evidence from Weakest Link," *Journal of Law and Economics*, 47, 431-52, 2004.

List, J., "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *The Quarterly Journal of Economics*, 2004, 119(1), 49-89.

List, J., "Friend or Foe? A Natural Experiment of the Prisoner's Dilemma," *Review of Economics and Statistics*, forthcoming, 2006.

Manski, C., "Measuring Expectations," *Econometrica*, 72(5), 1329-76, 2004.

Mobius, M. and T. Rosenblatt, "Why Beauty Matters," forthcoming *American Economic Review*, 2006.

Moreno, A., H. Nopo, J. Saavedra and M. Torero, "Gender and Racial Discrimination in Hiring: A Pseudo Audit Study for Three Selected Occupations in Metropolitan Lima," IZA Working Paper No. 979, 2004.

Ñopo, H., J. Saavedra and T., Maximo; (2005). "Ethnicity and Earnings in Urban Peru," forthcoming *Economic Development and Cultural Change* and previously IZA DP No. 980, 2003.

Peruvian Truth and Reconciliation Commission, “Final Report,” available at http://www.cverdad.org.pe/ingles/ifinal/conclusiones.php, 2003.

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

Riach, P. and J. Rich, “Field Experiments of Discrimination in the Market Place,” *Economic Journal,* CXII, F480–F518, 2002.

Torero, M., J. Saavedra, H. Nopo and J. Escobal; (2004). “An Invisible Wall? The Economics of Social Exclusion in Peru,” *Social Inclusion and Economic Development in Latin America*. Mayra Buvinic, Jaqueline Mazza and Ruthane Deutsch (Eds.) IDB and Johns Hopkins University Press.

Walker, C., *Smoldering Ashes: Cuzco and the Creation of Republican Peru, 1780-1840*, Durham: Duke University Press, 1999.

Appendix

Fixed-Effects Regression on Ranking (1=Highest, 19=Lowest)		
Variables	**Contribution & Photo**	**Two Types**
Age (years)	0.03 (0.30)	-0.02 (0.69)
Male	0.03 (0.90)	0.10 (0.76)
White	-0.19 (0.59)	0.16 (0.78)
Indigenous	-0.14 (0.70)	-0.32 (0.52)
Attractive	0.58 (0.47)	0.83 (0.39)
Unattractive	0.99 (0.23)	1.21 (0.18)
Expected Rank	0.85 (0.00)	0.72 (0.00)
Expected Rank*Attractive	-0.05 (0.52)	-0.16 (0.04)
Expected Rank*Unattractive	-0.03 (0.70)	-0.06 (0.47)
Constant	0.34 (0.76)	3.20 (0.01)
Individual Fixed Effects	*Yes*	*Yes*
Within-R2	0.71	0.45
N	760	760
p-values in parentheses		