



INTER-AMERICAN DEVELOPMENT BANK  
BANCO INTERAMERICANO DE DESARROLLO  
LATIN AMERICAN RESEARCH NETWORK  
RED DE CENTROS DE INVESTIGACIÓN  
RESEARCH NETWORK WORKING PAPER #R-543

**ETHNIC AND SOCIAL BARRIERS  
TO COOPERATION: EXPERIMENTS STUDYING  
THE EXTENT AND NATURE  
OF DISCRIMINATION IN URBAN PERU**

BY

MARCO CASTILLO\*  
RAGAN PETRIE\*\*  
MÁXIMO TORERO\*\*\*

\*GEORGIA INSTITUTE OF TECHNOLOGY

\*\*GEORGIA STATE UNIVERSITY

\*\*\*INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

APRIL 2008

**Cataloging-in-Publication data provided by the  
Inter-American Development Bank  
Felipe Herrera Library**

Castillo, Marco.

Ethnic and social barriers to cooperation : experiments studying the extent and nature of discrimination in urban Peru / by Marco Castillo, Ragan Petrie, Máximo Torero.

p. cm.  
(Research Network Working papers ; R-543)  
Includes bibliographical references.

1. Marginality, Social—Peru. 2. Sex discrimination against women—Peru. 3. Race discrimination—Peru. 4. Peru—Race relations. I. Petrie, Ragan. II. Torero Cullen, Máximo. III. Inter-American Development Bank. Research Dept. IV. Latin American Research Network. V. Title. VI. Series.

HN350.Z9 .C44 2007  
303.60985 C44----dc22

©2008  
Inter-American Development Bank  
1300 New York Avenue, N.W.  
Washington, DC 20577

The views and interpretations in this document are those of the authors and should not be attributed to the Inter-American Development Bank, or to any individual acting on its behalf.

This paper may be freely reproduced provided credit is given to the Research Department, Inter-American Development Bank.

The Research Department (RES) produces a quarterly newsletter, *IDEA (Ideas for Development in the Americas)*, as well as working papers and books on diverse economic issues. To obtain a complete list of RES publications, and read or download them please visit our web site at: <http://www.iadb.org/res>

## Abstract\*

This paper presents a series of experiments on discrimination in urban Lima, Peru. The experiments exploit degrees of information on performance as a way to assess how personal characteristics affect how people sort into groups, and the results show that behavior is not correlated with personal socio-economic and racial characteristics. However, people do use personal characteristics to sort themselves into groups. Height is a robust predictor of being desirable, as is being a woman. Looking indigenous makes one less desirable, and looking “white” increases one’s desirability. Interestingly, our experiments show that once information on performance is provided, almost all evidence of discrimination is eliminated. Although there is evidence of stereotyping or preference-based discrimination, clear information trumps discrimination.

---

\* This paper was undertaken as part of the Latin American and Caribbean Research Network Project “Discrimination and Economic Outcomes.” The authors would like to thank the Inter-American Development Bank for funding. We also thank Kevin Ackaramongkolrotn, Jorge de la Roca, David Solis, and Néstor Valdivia.

## 1. Introduction

Trust plays an important role in our choice of personal interactions. Trust is reflected in where we choose to live, whom we choose to befriend, and the groups to which we belong. While many choices are made with information on the qualities or reputations of others, some choices may be made with little more information than the impressions we form by driving through a neighborhood or viewing the clientele of a store. Lack of information can therefore prove a hindrance to economic exchange if people have misperceptions of the trustworthiness of others. People may withdraw from or never enter into interactions with certain segments of the population because of superficial perceptions, and initial perceptions might persist even in the face of evidence contradicting them. In the long run, society may suffer persistent losses due to exclusion if enough sorting takes place.

How important are these types of misperceptions in determining group composition and therefore economic outcomes? In this study, we want to know the salience of both performance and observable characteristics in how people sort into groups. We conjecture that people use observable characteristics, such as gender or race, to choose group members because they lack better information on future performance. However, even if people use personal characteristics only as a way to gauge information, if performance and characteristics are highly correlated, then we cannot tease apart which of the two is most salient in group membership. We use a series of experiments that break this correlation and allow us to assess which of the two criteria, personal characteristics or performance, is more salient. More importantly, we use a cross-section of the population as a way to reach a more diverse subject population than is normally found in standard laboratory experiments with college students.

Discrimination and social exclusion in the form of racial or ethnic discrimination seems to be critical in a multiracial and multilingual country as Peru, where indigenous groups and ethnic minorities are more likely to be poor than other groups. Previous work had shown that social exclusion in access to different markets—labor, credit, access to education—is a crucial issue in Peru. Discrimination and exclusion related to ethnicity, culture, physical appearance and religion take place in ways both obvious and subtle. Moreover, and as shown by Castillo and Petrie (2007), using data collected from the Peruvian Truth and Reconciliation commission, some patterns of human rights violations are difficult to reconcile with theories of statistical discrimination. If exclusion in Peru combines statistical as well as preference-based

discrimination, it is important to identify the extent of each and devise institutions that diminish both.

Group membership may have important economic benefits, such as the benefits from belonging to trade associations or investment groups. If the composition of the group dictates the benefits, then we may need to be careful whom we choose to be in our group or to which group we choose to belong; such sorting could have important consequences for which groups do well economically and which groups do not. If certain groups of people are unfortunate enough to have, for example, weak social networks and are perceived as having an untrustworthy appearance, they may be excluded from groups and only be able to find membership in low performing groups. Also, people conscious of discrimination might exclude themselves from groups as a way to avoid being discriminated against.

We use repeated linear public goods game experiments to explore these issues. Repeated public goods experiments are a natural environment in which to study trust, as they offer participants an opportunity to engage in reciprocal behavior. Level of cooperation, or reciprocity, has been found to depend on the initial propensities people in the group have to cooperate (Andreoni and Petrie, 2006). People will therefore sort themselves into groups of high performers. If people are not altruistic, then trust becomes important in this environment. Without trust in others' willingness to contribute to the public good, social benefits will not be achieved.

Since identification of discrimination for other than statistical reasons requires breaking the correlation between actions and appearances, we use several experimental treatments that manipulate the correlation between behavior and appearances. Subjects are shown digital photographs of others in the experiment and information on past performance and then asked to choose who they would like to have in their group. Our approach is novel in that it manipulates the equilibrium at the experiment level to identify sources of discrimination. A policy implication of this study is therefore to identify the changes in incentives necessary to reduce the prevalence of discrimination.

Our results show that people discriminate based on appearance and socio-economic characteristics despite the fact that there is no correlation between those characteristics and performance. That is, discrimination in urban Lima cannot be reconciled with theories of statistical discrimination. While the evidence is consistent with the presence of stereotyping or

taste-based discrimination, we also show that providing information on previous performance makes evidence of discrimination disappear almost completely. While this is encouraging, the results also show that there is evidence of preference-based discrimination since stereotyping is no longer a reasonable explanation once information on performance is revealed.

## **2. Appearance and Information**

Why might we think that appearance and information will interact to affect decisions? Previous research supports the notion that the social context of decisions can affect outcomes. Research in experimental economics has shown that being able to identify one's partner increases levels of altruism in dictator games (Bohnet and Frey, 1999; Burnham, 2003) and that cooperation in public goods games is increased by combining identification and information on past actions (Andreoni and Petrie, 2006). Also, people may have mistaken perceptions of behavior, expecting women to be more trusting than they actually are (Petrie, 2004).

Identification alone can increase cooperation, but specific characteristics of a partner, such as gender and beauty, can affect decisions. People are more cooperative and trusting with attractive people (Andreoni and Petrie, 2006; Petrie, 2004; and Eckel and Wilson, 2002), and attractive people make more money (Mobius and Rosenblatt, 2005; Hammermesh and Biddle, 1994). Decisions are also affected by the ethnic composition (Cummings and Ferraro, 2003) and the gender and age composition of the experimental group (Carter and Castillo, 2003).

Sorting or preference for individuals with certain observable characteristics may be a reflection of preference based or statistical based discrimination. Previous research using audit studies and field experiments have shown that there is evidence for both. Audit studies suggest findings consistent with preference-based discrimination (Riach and Rich, 2002), but List (2004) suggests that audit studies cannot distinguish this from statistical discrimination. List uses a sequence of field experiments at a sports card market to show that differentiated behavior is more likely due to statistical discrimination than pure discrimination.

Our experiments exploit a standard public goods game and vary one parameter to evaluate whether sorting is due to statistical or preference-based discrimination. This approach is attractive because it relies only on the comparative static properties of the equilibrium with discrimination.

### **3. Theoretical Motivation**

Standard economic reasoning implies that the way people sort into groups reveals their incentive to form coalitions. People will sort into the groups that maximize expected future gains, and observable characteristics of participants are important insofar as they reveal information on likely strategies to be played. In equilibrium, people will play their best responses to the expected behavior of others and others' expectations of one's behavior. This means that people will adjust their behavior according to their beliefs of what others are likely to do.

Observable characteristics are likely to be more salient and affect play in games in the absence of information on the likely play of others. This is the basis of statistical discrimination. Also, behavior towards others might be due to preferences for or against certain others regardless of beliefs. If people have preferences for the composition of the group, how people sort into groups will no longer solely reflect the incentive to maximize expected future gain.

Since one's quality as a partner is private information, there might be incentives to signal quality or to obtain information on the quality of others, and people would have an incentive to form reputations. In order to avoid any reputation effects we need to eliminate the incentive to form a reputation in early rounds of the game.

This suggests a natural test on theories explaining sorting into groups. Statistical discrimination theories suggest that appearance affects sorting only because appearance provides information on expected behavior. Once information on behavior is provided, the role of appearance must be muted. But what if behavior is correlated with appearances? For example, what if Caucasians are indeed more cooperative? If this is the case, then we cannot determine if sorting along social characteristics in the presence of information on past behavior is evidence of pure discrimination or statistical discrimination. This identification problem can be resolved if this correlation can be broken, so that any subsequent sorting along social characteristics is due to pure discrimination. Our experimental design will allow us to observe if people engage in statistical discrimination or pure discrimination when choosing groups.

## **4. The Sample**

The proposed site for our experiments was urban metropolitan Lima in Peru. This site lends itself to Internet-based experiments that draw from a larger population because Internet cabins are common in Lima and a high proportion of the population with expertise using the internet. According to a survey conducted in 2003, there were 476 internet cabins distributed across all districts of Lima. This amounts to around 1 computer/hour per 10 people (assuming 10 computers per cabin, 12 hours of service and an urban population of 5,681,941 according to the census of 1993). This characteristic of urban Lima allows us to conduct internet-based experiments with non-college student populations—an important distinction, given that students belong to a potentially highly unrepresentative segment of the population, thus reducing the external validity of the results and preventing us from drawing clear policy implications.

### ***4.1 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 now do so. Second, at the same time we want to have a sample of subjects who were representative of the young working population in metropolitan Lima. With these objectives in mind we implemented a four-step strategy.

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

Secondly, based on these criteria we contracted two specialized companies specializing in surveys and recruiting. These companies' main assets is that they collect representative databases for metropolitan Lima and use local leaders to help in building the list of individuals in their databases, which helped to secure a diverse population in the experiments. In general, this mechanism ensures that the opportunity to participate in the experiment is distributed equally. From these databases we sampled all the potential subjects that met all of our criteria. From the resulting sub-sample we performed a random lottery and selected the individuals to be part of the experiment.



Thirdly, in the interest of over-sampling clusters of owners of small, medium and micro-enterprises, we send a team of recruiters to Gamarra (an industrial area in metropolitan Lima). The advantage of Gamarra was two-fold. First, we had a pre-census of all the establishments in Gamarra, which allowed us to randomly select buildings from which we invited subjects. Second, it is one of the biggest clusters of small and medium enterprises (SMEs) in metropolitan Lima and represents a rich mix of population in terms of place of origin and socio-economic background. According to the census of 1993, 37 percent of the population of La Victoria (where Gamarra is located) consisted of migrants. Moreover, the patterns of migration are diverse, allowing us to sample from a diverse population.

The procedure followed to select the establishments within Gamarra consisted of the following protocol. First, based on the pre-census, we identified the blocks with the highest concentration of commercial buildings, which consisted of three long blocks (approximately 8 standard 100mt blocks) crossed by three transversal blocks. Second, we visited four commercial buildings per block, randomly selecting one building on that block and three buildings to its left or right. Third, within the buildings we selected establishments by performing a random lottery from the total number of establishments in each odd-numbered level of the building. Fourth, we invited one person in the establishment to participate in the experiment.

Finally, to be representative of subjects of university age and that are part of the university system, we also invited potential subjects that were taking classes at extension institutes within the major universities located in metropolitan Lima, which have student bodies that represent very diverse socio-economic backgrounds. In addition, this group of students from the extension institutes is different from college students in the sense that they are already working and do not necessarily have a college degree. Recruitment of subjects in all of these sites followed standard experimental recruitment protocols.

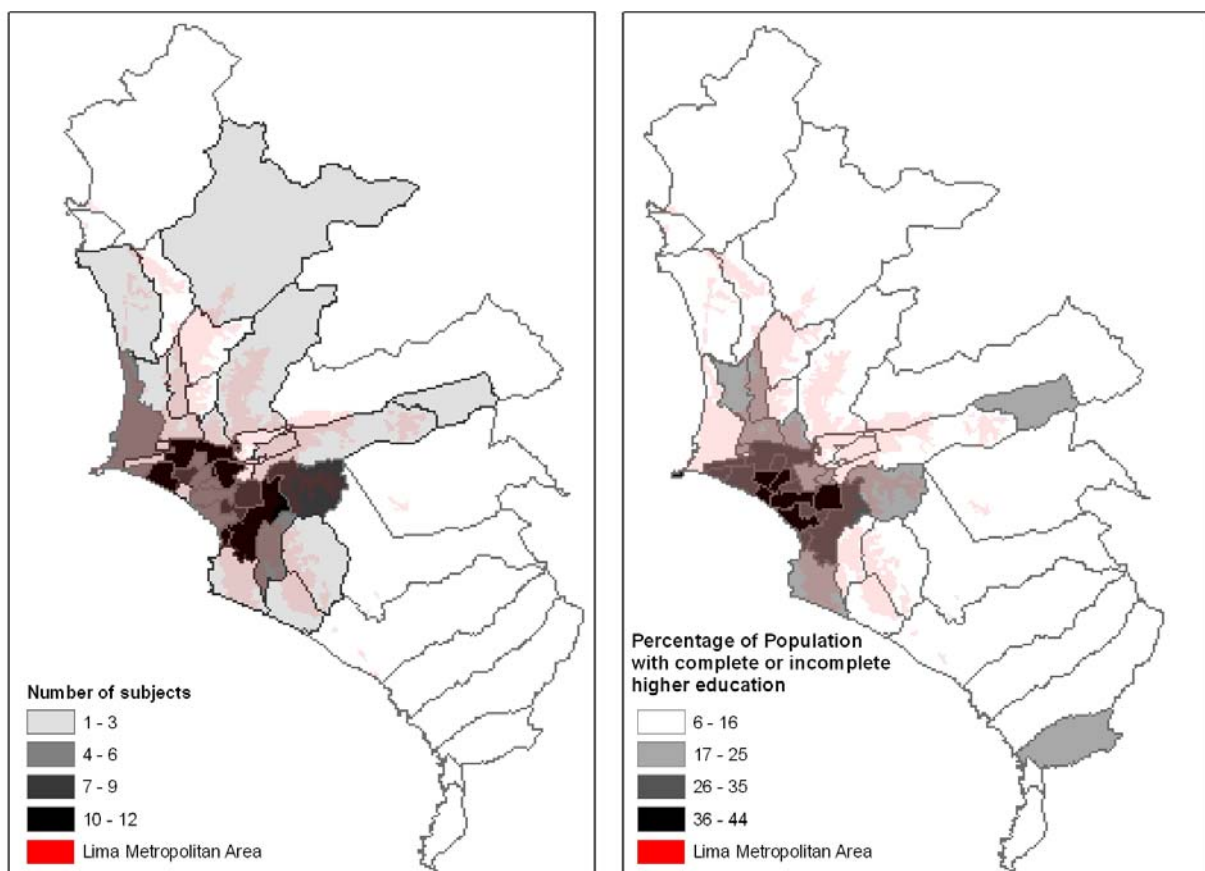
It is important to note that because our experiments rely on Internet protocols and the knowledge of using a computer, we will likely exclude some segments of the population that might suffer more marked patterns of discrimination. Therefore it is important to recognize that our initial results will therefore give a lower bound estimate to the extent of discrimination. Even so, the Internet protocols used were simple enough as to include large segments of the population. Previous experience by the researchers in rural areas in South Africa and Central America show that illiterate subjects are able to understand experimental procedures presented in

a graphical manner. The experiments in this research required simply that the subjects know how to use a computer mouse.

#### ***4.2 Comparison of Sample to Metropolitan Lima Population***

Map 1 give a first 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.

**Map 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 representativeness of our sample, we used the 2004 *Encuesta Nacional de Hogares* (ENAHO) and compiled from it a sub-sample that meets the same eligibility criteria as 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 sample and ENAHO 2004 for a subset of variables common to our survey and to the ENAHO. Both samples have a similar distribution among almost all the variables, although our sample is slightly more educated and had significantly fewer migrants than the ENAHO. This is most likely a reflection of the requirement in our sample that subjects know how to use the Internet. Nonetheless, the age, gender, monthly income, average education, household size, and characteristics of the household are very similar. This comparison gives us confidence that our sample is a reasonable representation of the larger population in metropolitan Lima.

**Table 1. Comparing Our Sample with ENAHO 2004**

	Sample	ENAHO 04 <sup>1/.</sup>
Age	26.31	27.95
Male	61.6%	54.6%
Level of education <sup>2/.</sup>		
Incomplete non-university tertiary	14.6%	24.6%
Complete non-university tertiary	17.5%	36.2%
Incomplete university tertiary	35.0%	20.0%
Complete university tertiary	32.8%	19.2%
Years of education <sup>2/.</sup>	14.61	14.02
Native language <sup>3/.</sup>		
Quechua	0.6%	1.5%
Aymara		
Spanish	95.0%	98.5%
Other	4.4%	
Migrant <sup>4/.</sup>	15.2%	37.9%
Weekly hours of work <sup>5/.</sup>	35.4	44.7
Monthly income <sup>5/.</sup> <sup>6/.</sup>	1087.3	1099.2
Health insurance <sup>7/.</sup>	63.3%	40.3%
Affiliation to pension fund	37.6%	50.8%
Members of household	4.9	5.1
15 years or younger	0.7	0.9
65 year or older	0.4	0.3
Working people in household	2.50	2.54
Rooms in dwelling		
One room	3.2%	4.0%
Two rooms	10.1%	8.7%
Three rooms	27.2%	19.1%
Four rooms	17.7%	25.0%
Five rooms	15.2%	15.2%
Six rooms	10.8%	10.2%
Seven or more rooms	15.8%	17.7%

1/. Estimations for ENAHO 2004 have been limited to observations with the following characteri  
(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 be  
excluded from schooling-related estimations.

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

4/. Migrants are defined as those not born in Lima.

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

6/. In constant soles of May 2006.

7/. EsSalud or private health insurance.

The sample of individuals for this study was composed of self-employed or dependent workers from the high, middle and lower-income classes of metropolitan Lima. They were relatively young, 23 years old on average, and most had some kind of higher education. The average number of years of schooling was 14.6, with a standard deviation of 1.9, and close to 70 percent of the sample had complete or incomplete university studies. Only 20 percent of individuals graduated from a public school, while 45 percent graduated from a private religious high school. Although 45 percent of subjects' fathers and 27 percent of their mothers had finished university studies, their parents' education was on average lower than their own. All subjects had some labor experience, and 98 percent had worked during the last 12 months. More than two-thirds of the applicants were males. Family and per capita income were generally higher for those individuals who attended high school at a private institution, had undertaken technical or professional studies at a university and whose parents had achieved post-secondary diplomas, whether at universities or vocational institutions.

With regard to asset ownership, we found that most subjects live in dwellings with more than two rooms, and 40 percent of the sample lives in dwellings of five or more rooms. Although the sample of applicants for our study was drawn from a specific segment of the population of Metropolitan Lima that had relatively homogeneous schooling, experience and age. Nonetheless, we found a high dispersion in earnings among the subjects and their households, which assured that they belonged to different socioeconomic levels.

## 5. Experimental Design

We use a linear public goods game to explore discrimination in group formation. 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 *céntimos* to the subject.<sup>1</sup> Each token placed in the public investment yields a return of  $\alpha_i$  to the subject and every other member of the group; the return to the public investment varies across treatments and will be explained later. 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,  $\pi_i$ , and the total number of tokens contributed to the public investment by the group,  $G$ . In total, subjects play three 10-round

---

<sup>1</sup> There are 100 *céntimos* in 1 *sol* (the Peruvian currency). At the time of the study, US\$1 = 3.2 *soles*.

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 their preferences, we created the following procedure. Subjects rank all other subjects from most preferred to least preferred, and 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 those 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 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 person's 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. 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 James prefers Jill's company to Jane's company, then James always prefers a group that exchanges Jane for Jill, regardless of who the other members of the group are. Under these conditions, revealing the ordering of others is a weakly dominant strategy for James. If James 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 are additive, it does not matter whether he is chosen first or last.

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

The mechanism described above is incentive compatible, regardless of the preferences over groups, if people are able to rank all possible groups that one could be paired with. Unfortunately, this option would be impractical since the number of groups to be ranked would be exceedingly large. 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.

There are four experimental treatments: Contribution Only, Photo Only, Contribution and Photo, and Two Types. Treatments differ in the  $\alpha_i$  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  $\alpha_i = 2$  *céntimos*, 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 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, while 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,  $\alpha_i \in \{0.5, 5.0\}$  *céntimos*. Half of the subjects are randomly assigned a value of 0.5 and half are randomly assigned a value of 5.0, and 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. 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 value to invest little to nothing in the public good.

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-questionnaire. Subjects were recruited from diverse neighborhoods in Lima, Peru. The experiments were conducted on computers in two computer labs at the University of the Pacific in Lima, Peru. Since most subjects worked full time, the experiments were conducted on weekend afternoons. Sixty percent of the subjects are men, and subjects ranged in age from 20-30 years old, with an average age of 26.3 years (standard deviation 4.2 years). More descriptive statistics on the subjects are in the Results section. Race, beauty and height classifications are discussed in the following section.

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).

## **6. Race, Beauty and Height 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 and beauty are 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 subjects in the experiment, to rate the photos of the subjects in terms of race and beauty, as well as height. A rater only rated the photo in terms of one characteristic, race, beauty or height, and not all three.

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 (2004). 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, which are readily recognized as distinct racial groups. This gives a more nuanced measure of race and more accurately captures racial mixing in Peru.

To obtain these ratings, we had 20 non-participants in the public goods experiment (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 10, with zero being complete



absence of the dimension and 10 being the most intense. Raters were instructed to choose whichever number between one and 10 best 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 \$9.67 for their time.

For the beauty ratings, 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 each photo in terms of physical attractiveness. The scale went from one to nine, with one being not at all attractive and nine being very attractive. Raters were asked to choose the number that best described that person's physical attractiveness to them.

For the height guess, we followed the same procedure as with the race and beauty ratings. The only difference is that the 10 men and 10 women were asked to guess the height, in centimeters, of each person in the photo. Raters were free to choose any number for the height.

In terms of agreement among raters, there was usually 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.<sup>2</sup>

While the rating scale ranged from zero to ten for race and one to nine for beauty, 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 six. 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, beauty or height to obtain the final ratings we use to analyze the data.

---

<sup>2</sup> The Cronback alpha for interrater reliability is another measure of agreement among raters. For the white dimension, the coefficient is 0.9565; for the indigenous dimension, the coefficient is 0.9285; for the black dimension, the coefficient is 0.9451; and for the Asian dimension, the coefficient is 0.9113.

For race, the most likely intensities in the subject population are white and indigenous. 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 percent of the population and Asians make up 3 percent 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 identified themselves primarily as a mix of white and indigenous, we concentrate on these two dimensions in our analysis of contributions and ranking in the next section.

## **6. Results and Discussion**

Table 2 present descriptive statistics of the experimental subjects.<sup>3</sup> Three out of five subjects are men, and the average age is 26 years. As mentioned in the sample description section, our sample is slightly more educated than the population at large. On average, participants have three or more years of post-secondary education, and 29 percent have a college degree. The sample also reflects the different ethnic and cultural background of Lima's population. Sixteen percent of the sample has at least one grandparent whose mother tongue was neither Spanish nor any other Peruvian indigenous languages. Thirty percent of the sample has at least one parent whose mother tongue was indigenous to Peru. While stature is a self-reported variable, we find great variation in height. On average, a male subject reports being 1.73 meters tall and a female subject reports being 1.63 meters tall. Finally, experimental subjects live in household with an average of five persons.

---

<sup>3</sup> There are three post-experiment surveys missing from the sample

**Table 2. Descriptive Statistics**

Variables	N	Mean	S.D.	Min	Max
1[Men]	160	0.61	0.49	0	1
Age (years)	160	26.28	4.23	20	35
Education (years)	160	15.07	1.72	10	19
1[college degree]	160	0.29	0.46	0	1
1[inc. college degree]	160	0.32	0.47	0	1
European Grandparents (n)	157 <sup>1</sup>	0.17	0.60	0	4
Indigenous Grandparents (n)	157 <sup>1</sup>	0.31	0.89	0	4
Height (meters)	160	1.69	0.08	1.52	1.94
Family Size	156 <sup>1</sup>	4.95	2.24	1	13
1[Religious High School]	160	0.44	0.50	0	1

<sup>1</sup> There are three missing post-experiment questionnaires from the sample.

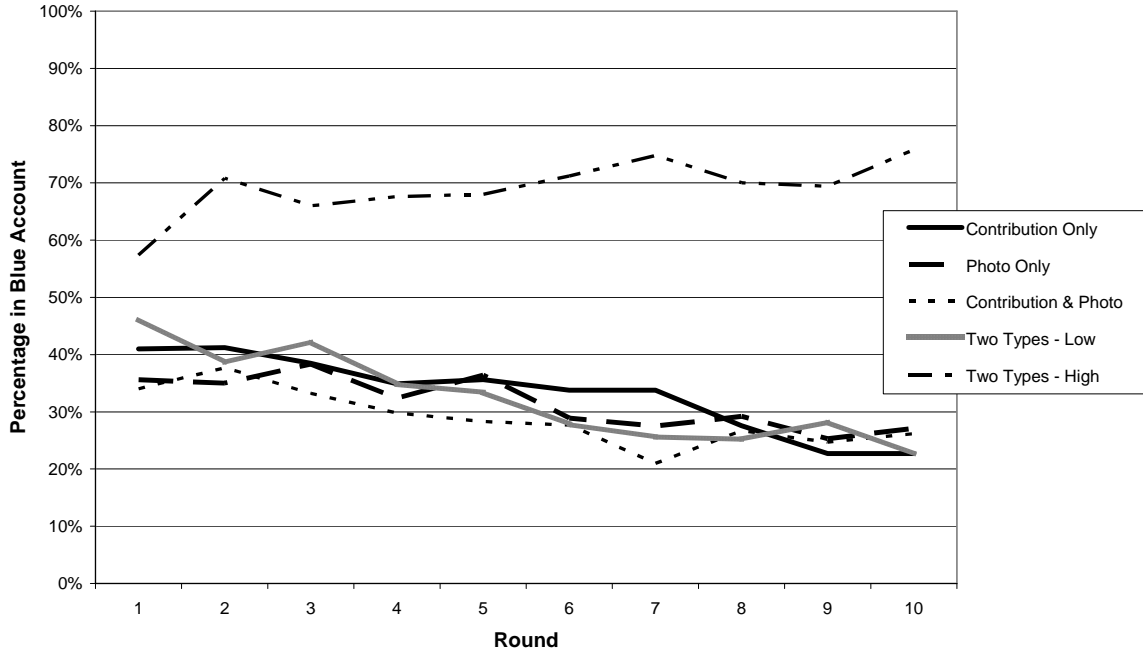
### **6.1 What Did People Do in the Experiment?**

Figures 1 and 2 show the aggregate behavior in all experimental sessions. Across all rounds of the first sequence of the experiment, contributions to the public good range from 23 percent of subjects' endowment for low types in the Two Types treatment to 46 percent of subjects' endowments in Contribution Only. As it is commonly observed (see Kagel and Roth, 1995), contributions tend to decline with time. Contributions decline to 9 percent for low types in the Two Types treatment and to 27 percent in Contribution Only. A similar pattern is observed in the second sequence of the experiment in Figure 2. Contributions in the first round of the second sequence of the experiment range from 30 percent in Contribution Only to 46 percent for low types in Two Types. Contributions in the last round of the second sequence decrease to 14 percent in Photo Only and to 32 percent for low types in Two Types.

The figures also show that the incentives of the Two Types treatment successfully induce a separation in behavior between high and low types. High types contribute 50 percentage points more to the public good than low types. Moreover, the figures also show convergence towards play of dominant strategies by high types.

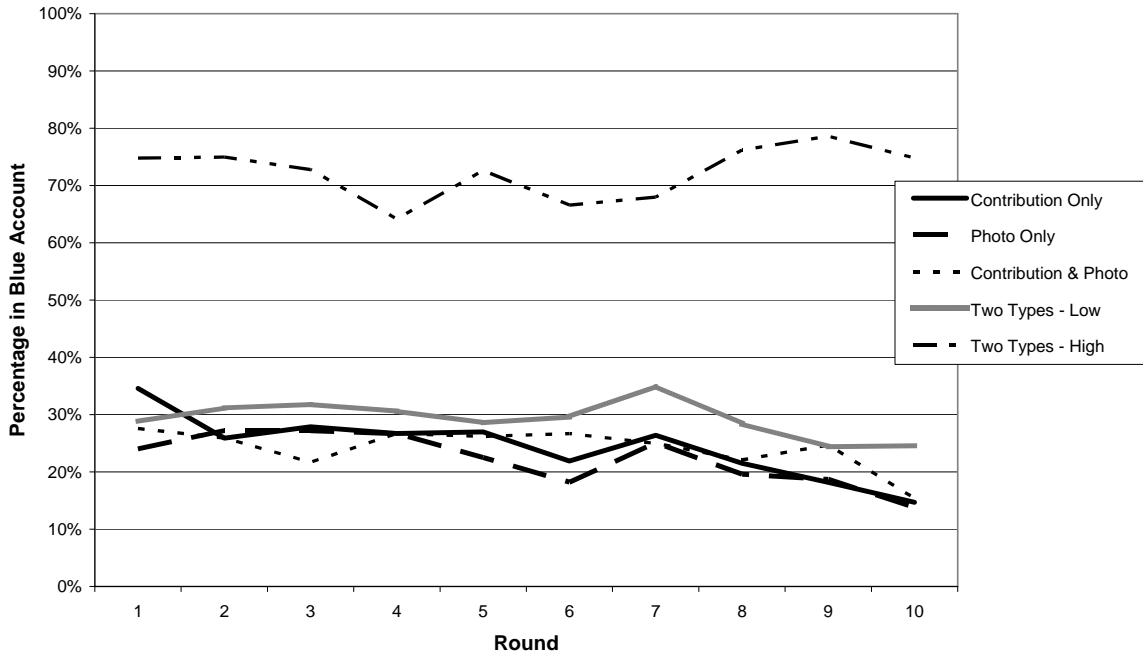
**Figure 1.**

**Contributions to the Public Good  
First Sequence**



**Figure 2.**

**Contributions to the Public Good  
Second Sequence**



A basic premise in theories of statistical discrimination is that people of different backgrounds might behave differently and therefore, in the absence of better information, ethnic or cultural background can be used as a proxy of behavior. For instance, migrants might experience unfavorable market conditions, causing them to behave selfishly. Conversely, more affluent subjects can afford to be more altruistic or take more risks. Table 3 shows a series of 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.<sup>4</sup>

**Table 3. Percent of Endowment Contributed to the Public Good  
(Sequence 2)**

Variables	Contribution Only, Photo Only, and Contribution and Photo	Two Types	All Treatments
1[Men]	4.62 (0.21)	-3.03 (0.75)	3.19 (0.37)
Age (years)	0.10 (0.78)	-0.81 (0.40)	-0.13 (0.70)
Education (years)	0.55 (0.56)	-2.74 (0.14)	-0.57 (0.51)
Height (meters)	-1.56 (0.95)	117.36 (0.03)	27.78 (0.21)
1[White>Med, Indig.≤ Med]	-0.16 (0.96)	-2.90 (0.74)	-0.96 (0.77)
1[White≤Med, Indig.> Med]	-1.55 (0.68)	4.09 (0.70)	-0.30 (0.94)
1[Religious High School]	-1.72 (0.58)	5.99 (0.48)	1.16 (0.70)
1[Low Type]	-	-	-13.44 (0.72)
1[High Type]	-	48.01 (0.00)	30.37 (0.42)
Round	-1.19 (0.00)	-0.12 (0.73)	-0.93 (0.00)
Constant	26.49 (0.51)	-113.41 (0.19)	-
<i>Individual Random Effects</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
<i>Group Fixed Effects</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Within-R2	0.0360	0.0003	0.0209
N	1200	400	1600
p-values in parentheses			

<sup>4</sup> Results are robust to different specifications.

The regressions in Table 3 show that behavior is not correlated with personal characteristics. On average, contributions decrease by 10 percent from round 1 to round 10. There is a slight effect of taller people giving more in the Two Types treatment. It is further interesting to compare 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. The regressions show that even when the relative cost of contributing was higher for low types in the Two Types treatment than in other treatments they contributed 27 percentage points more. Since high types in the Two Types treatment contribute 71 percentage points more than in other treatments, this evidence is consistent with theories of reciprocity and peer effects.

Table 3 shows that personal characteristics are of little help in predicting others' behavior. This result will be useful in interpreting the results in the following section. It is important to note that ethnic background, as measure as intensity of a racial characteristic, is not correlated with behavior at all.

## ***6.2 How Were People Ranked?***

The previous section shows that there is little evidence supporting the hypothesis that personal characteristics correlate with behavior. This section investigates if personal characteristics are used when choosing groups. The regression is based on a few covariates due to the fact that results are not altered significantly by the addition of additional ones. Ethnicity is measured by the average standardized intensity variable of white and indigenous described in the section on Race Classification. We use these aggregated racial intensities to create a discrete variable determining if a person is white or indigenous. A person is considered white if her average of racial intensity ratings in the white dimension is above the median and her average racial intensity ratings in the indigenous dimension is below the median. A person is considered indigenous if her average racial intensity in the indigenous dimension is above the median and her average racial intensity in the white dimension is below the median.

Table 4 reports the OLS regression for rankings separately for each treatment.<sup>5</sup> The dependent variable is the rank that a person is given. That is, a person with a rank of 1 is rank highest and a person with a rank of 19 is rank lowest. Given the way rank is defined, the

---

<sup>5</sup> The results in Table 4 and Table 5 are similar if using rank-ordered logit or if we use robust standard errors. Note that the reported results do not use robust standard errors. The results are also similar if we use racial intensities of trained raters.

interpretation of the sign of coefficients must be adjusted accordingly. 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.

There are two covariates that require 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. The expected coefficient on this variable should be 1 if information on others' behavior is the only relevant information in creating ranks.

**Table 4. OLS Regression on Individual Ranking  
(highest = 1, lowest = 19)**

Variables	Photo Only	Contribution & Photo	Two Types
Age (years)	0.06 (0.28)	0.03 (0.32)	-0.02 (0.67)
1[Man]	2.89 (0.00)	0.09 (0.81)	-0.00 (0.99)
Height (meters)	-10.37 (0.00)	-0.85 (0.65)	-1.10 (0.66)
Expected Rank	-	0.83 (0.00)	0.64 (0.00)
1[White > Med & Indig. ≤ Med]	0.19 (0.73)	-0.06 (0.86)	-0.71 (0.14)
1[White ≤ Med & Indig. > Med]	1.47 (0.00)	-0.34 (0.28)	-0.19 (0.68)
Constant	23.48 (0.00)	2.30 (0.46)	6.17 (0.13)
R2	0.05	0.70	0.43
N	760	760	760

People seem to understand that having high contributors in the group is the best strategy. For instance, Expected Rank alone explains 67 percent of the variance of ranks in Contribution Only (not shown in Table 4). Expected Rank remains a strong predictor of rank in all treatments where information on previous contribution is provided.

Interestingly, despite the fact that personal characteristics have no bearing on what people do in the experiment, they tend to predict the way people are ranked. In the Photo Only treatment, men are ranked on average 2.4 ranks lower than women. Height also has a strong effect on the way people are ranked: 10 extra centimeters of height increases rank by one. Tall women are therefore ranked rather high. Due to the fact that people only saw the picture of other participants, the result on height is puzzling. Height might be correlated with other characteristics captured in a photo and might therefore not measure the impact of height per se. However, as mentioned before, we collected data from independent people to check if people we

able to correctly guess the height of others by looking at head-shot pictures. Indeed, the average guessed height reported by independent raters is highly correlated with real height even after controlling for sex and ethnicity. That is, we cannot discard the hypothesis that height itself explains how people are ranked.

Relevant for the question of racial discrimination, the regression on rankings made in Photo Only also shows that people that look indigenous are ranked 1.47 ranks lower. Table 4, however, shows that discrimination based on race is present only when no information on past performance is available. Rankings made in the treatment with Contributions and Photos show that race indicators are no longer significant. That is, the regressions are consistent with stereotyping but not preference-based discrimination.

**Table 5. OLS Regression on Individual Ranking  
(highest = 1, lowest = 19)**

Variables	Photo Only		Contribution & Photo		Two Types	
	Men	Women	Men	Women	Men	Women
Age (years)	0.06 (0.31)	0.04 (0.63)	0.01 (0.68)	0.08 (0.22)	0.07 (0.21)	-0.10 (0.03)
1[Man]	2.76 (0.00)	3.12 (0.00)	0.05 (0.89)	0.07 (0.93)	0.26 (0.69)	-0.25 (0.62)
Height (meters)	-9.17 (0.03)	-12.56 (0.03)	-1.08 (0.75)	0.41 (0.92)	-2.95 (0.46)	0.79 (0.80)
Expected Rank	-	-	0.86 (0.00)	0.73 (0.00)	0.55 (0.00)	0.73 (0.00)
1[White > Med & Indig. ≤ Med]	0.39 (0.58)	-0.23 (0.81)	-0.11 (0.75)	0.12 (0.86)	-1.09 (0.16)	-0.41 (0.49)
1[White ≤ Med & Indig. > Med]	1.84 (0.00)	0.74 (0.42)	0.01 (0.97)	1.34 (0.06)	-0.42 (0.56)	-0.03 (0.96)
Constant	21.06 (0.00)	27.85 (0.00)	2.93 (0.39)	-0.53 (0.94)	7.96 (0.22)	4.27 (0.66)
R2	0.05	0.04	0.74	0.61	0.32	0.55
N	494	266	551	209	361	399

Who is doing the discriminating? Table 5 shows how men and women rank others. In Photo Only, both men and women rank tall women higher, but it is men who rank people who look indigenous lower. Women react more strongly to tall women than men. In Contribution and Photo, women rank people who look indigenous lower. Men do not react to racial characteristics.



**Table 6. OLS Regression on Individual Ranking  
(highest = 1, lowest = 19)**

Variables	Photo Only		Contribution & Photo		Two Types	
	White	Indigenous	White	Indigenous	White	Indigenous
Age (years)	0.03 (0.71)	0.13 (0.13)	0.02 (0.55)	-0.01 (0.87)	0.10 (0.04)	-0.09 (0.12)
1[Man]	4.08 (0.00)	2.58 (0.00)	0.73 (0.07)	-0.01 (0.99)	-0.12 (0.82)	0.40 (0.56)
Height (meters)	-10.89 (0.07)	-11.99 (0.03)	-2.07 (0.31)	-0.32 (0.93)	-2.60 (0.45)	1.71 (0.67)
Expected Rank	-	-	0.92 (0.00)	0.73 (0.00)	0.80 (0.00)	0.69 (0.00)
1[White > Med & Indig. ≤ Med]	-0.63 (0.51)	1.20 (0.18)	-0.35 (0.27)	0.25 (0.70)	-0.42 (0.52)	-1.39 (0.07)
1[White ≤ Med & Indig. > Med]	1.97 (0.03)	0.91 (0.28)	0.21 (0.56)	0.83 (0.19)	-0.09 (0.88)	-0.53 (0.48)
Constant	24.21 (0.00)	24.43 (0.00)	3.28 (0.33)	3.16 (0.62)	3.73 (0.50)	3.33 (0.66)
R2	0.12	0.03	0.90	0.54	0.68	0.48
N	247	304	228	266	228	266

Table 6 shows how White and Indigenous people rank others. Both groups rate tall women higher, but only Whites rate indigenous-looking people lower. When information on contributions is known, this is a strong predictor of rank. Whites do rank men lower in Contribution and Photo, and they also rank older people lower in Two Types. But this effect is rather small.

Tables 7 and Table 8 show OLS regressions that further investigate the presence of discrimination across treatments.<sup>6</sup> Table 7 shows a linear probability model on the likelihood of being in the top four of any list. As mentioned above being a man decreases the probability of being among the top four and height increases the probability of being among the top four. Table 8 shows the likelihood of being in the bottom 4 of any list. Both of these tables confirm previous results.

<sup>6</sup> The results in Tables 6 through 8 are similar if we use probit or robust standard errors.

**Table 7. Probability of Being in the Top 4  
(highest = 1, lowest = 19)**

Variables	Photo Only	Contribution & Photo	Two Types
Age (years)	-0.01 (0.14)	-0.00 (0.39)	0.00 (0.21)
1[Man]	-0.17 (0.00)	-0.02 (0.57)	0.05 (0.18)
Height (meters)	0.60 (0.02)	-0.19 (0.27)	-0.19 (0.32)
Expected to be in Group	-	0.68 (0.00)	0.52 (0.00)
1[White > Med & Indig. ≤ Med]	-0.00 (0.92)	0.00 (0.97)	0.15 (0.00)
1[White ≤ Med & Indig. > Med]	-0.11 (0.00)	0.02 (0.61)	0.04 (0.27)
Constant	-0.49 (0.22)	0.43 (0.14)	0.20 (0.50)
R2	0.04	0.52	0.34
N	760	760	760

**Table 8. Probability of Being in the Bottom 4  
(highest = 1, lowest = 19)**

Variables	Photo Only	Contribution & Photo	Two Types
Age (years)	-0.00 (0.72)	0.01 (0.01)	-0.01 (0.01)
1[Man]	0.14 (0.00)	-0.04 (0.24)	0.02 (0.58)
Height (meters)	-0.62 (0.02)	-0.03 (0.86)	-0.34 (0.12)
Expected to be in Group	-	0.74 (0.00)	0.48 (0.00)
1[White > Med & Indig. ≤ Med]	0.13 (0.00)	0.09 (0.00)	0.01 (0.72)
1[White ≤ Med & Indig. > Med]	0.15 (0.00)	0.12 (0.00)	0.08 (0.05)
Constant	1.10 (0.01)	-0.14 (0.59)	0.82 (0.01)
R2	0.03	0.63	0.28
N	760	760	760

Finally, note the results for the Two Types treatment. This treatment is interesting because subjects are induced to behave quite differently regardless of their looks or background. Despite this, we find that looking white increases the likelihood of being named among the top four. Also, looking indigenous increases the likelihood of being named among the bottom four.

## 7. Conclusions and Policy Implications

We present a series of experiments aimed at determining the nature of discrimination in urban Lima, Peru. Subjects play a linear public goods game and are allowed to sort into groups. 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. 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, including ethnicity and socio-economic standing. That is, our results show that there is little room for statistical theories of discrimination. However, our experiments show that people do use the personal characteristics of others when given the opportunity to choose partners. Our research shows evidence of preference-based discrimination or stereotyping. Interestingly, our experiments also show that evidence of discrimination or stereotyping vanishes almost completely once information on others' behavior is provided.

Nonetheless, subjects tend to prefer groups of tall people, women, and white-looking people. While evidence of discrimination is almost completely eliminated by revealing information on others' behavior, there is still evidence that race is an important factor even when information is revealed. Intriguingly, while tall women are preferred in the absence of information, they are less likely to be selected for the top ranks when information is revealed. The effect of race, however, seems to be constant. This effect even survives when subjects are given incentives that make their behavior orthogonal to their personal characteristics.

The fact that not everyone uses others' characteristics in ranking in the same way provides further evidence of stereotyping or taste-based discrimination. While there is agreement across genders and ethnicities that taller people and women are more desirable partners, the effect of race on rankings is basically explained by the behavior of men and white participants. Since our experiments show that discrimination can be erased when information on performance is available, we conclude that these results are expression of prejudice.

Our research has 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. While our experiments show that information on others' performance is quite useful in solving initial stereotypes, it is clear that in practical terms it is difficult to provide precise and reliable measures of one's performance. That is, it is not clear that policymakers have the tools to make signals clearer or to make measurement of performance in the work place better. It is also entirely possible that, while discrimination in the work place is diminished through public intervention, other avenues such as marriage or neighborhood sorting survive.

Overall, our research shows that carefully designed experiments can allow us to identify the nature of discrimination.

## References

- Andreoni, J., and R. Petrie. 2004. "Public Goods Experiments Without Confidentiality: A Glimpse Into Fund-Raising." *Journal of Public Economics* 88(7-8): 1605-1623.
- . 2006. "Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments." Experimental Economics Working Paper 2006-2002. Atlanta, United States: Georgia State University.
- Bohnet, I., and B. Frey. 1999. "The Sound of Silence in Prisoner's Dilemma and Dictator Games." *Journal of Economic Behavior and Organization* 38: 43-57.
- Bogomolnaia, A., and M.O. Jackson. 2002. "The Stability of Hedonic Coalition Structures." *Games and Economic Behavior* 38(2): 201-230.
- Burnham, T.C. 2003. "Engineering Altruism: A Theoretical and Experimental Investigation of Anonymity and Gift Giving." *Journal of Economic Behavior and Organization* 50: 133-144.
- Carter, M., and M. Castillo. 2003. "An Experimental Approach to Social Capital in South Africa." Agricultural and Applied Economics Staff Paper 448. Madison, United States: University of Wisconsin-Madison, 2003.
- Castillo, M., and R. Petrie. 2007. "Discrimination in the Workplace: Evidence from a Civil War in Peru." "Discrimination in the Workplace: Evidence from a Civil War in Peru." Andrew Young School of Policy Studies Research Paper 07-37. Atlanta, United States: Georgia State University.
- Castillo, M., and Ragan Petrie. 2006. "Discrimination in the Lab: Experiments Exploring the Impact of Performance and Appearance on Sorting." Andrew Young School of Policy Studies Research Paper 07-17 Atlanta, United States: Georgia State University.
- Cummings, R., and P. Ferraro. 2003. "Inter-Cultural Discrimination in the Ultimatum Game: Ethnic Bias and Statistical Discrimination." Environmental Policy and Experimental Laboratory Working Paper 2003-02. Atlanta, United States: Georgia State University.
- Eckel, C., and R. Wilson. 2002. "Conditional Trust: Sex, Race and Facial Expressions in a Trust Game." Working Paper. Houston, United States Rice University.  
[http://www.ruf.rice.edu/~rkw/RKW\\_FOLDER/PC\\_2002\\_RKW.pdf](http://www.ruf.rice.edu/~rkw/RKW_FOLDER/PC_2002_RKW.pdf)
- Hammermesh, D., and J. Biddle. 1994. "Beauty and the Labor Market." *American Economic Review* 84(5): 1174-94.

- Kagel, J., and A. Roth. 1995. *Handbook of Experimental Economics*. Princeton, United States: Princeton University Press.
- List, J. 2004. “The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field.” *Quarterly Journal of Economics* 119(1): 49-89.
- Manski, C. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies* 60(3): 531-542.
- Mobius, M., and T. Rosenblatt. 2005. “Why Beauty Matters.” *American Economic Review* 96(1): 222-235.
- Ñopo, H., J. Saavedra and M. Torero. 2004. “Ethnicity and Earnings in Urban Peru.” IZA Discussion Paper 980. Bonn, Germany: Institute for the Study of Labor (IZA).
- Petrie, R. 2004. “Trusting Appearances and Reciprocating Looks: Experimental Evidence on Gender and Race Preferences.” Atlanta, United States: Georgia State University. Manuscript.
- Riach, P.A., and J. Rich. 2002. “Field Experiments of Discrimination in the Market Place.” *Economic Journal* 112: F480–F518.
- Torero, M. et al. 2004. “An Invisible Wall? The Economics of Social Exclusion in Peru.” In: M. Buvinic, J. Mazza and R. Deutsch, editors. *Social Inclusion and Economic Development in Latin America*. Washington, DC, United States: Inter-American Development Bank.