Ethnic and Social Barriers to Cooperation: Experiments Studying the Extent and Nature of Discrimination in Urban Peru

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

Ragan Petrie
Georgia State University
rpetrie@gsu.edu

Maximo Torero
IFPRI and GRADE (Lima, Peru)
m.torero@cgiar.org

September 6, 2006

Abstract

We present a series of experiments to understand the nature of 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. Our results show that behavior is not correlated with personal socio-economic and racial characteristics. That is, if discrimination exists in urban Lima, this cannot be explained by theories of statistical discrimination. However, our results show that 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. This is evidence of stereotyping rather than preference-based discrimination. Clear information trumps discrimination.

***** 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.

Introduction

Trust plays an important role in who we choose to interact with. It is reflected in where we choose to live, who we choose to be friend, and the groups we belong to. 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 be 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 perceptions. 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 research, we want to know how salient is performance and observable characteristics to 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 is most salient: personal characteristics or performance. More importantly, we use a cross-section of the population as a way to reach a more diverse subject population than 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. Indigenous or ethnic minorities are more likely to be poor than any other group. 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 (2005b), 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 who we choose to be in our group or which group we choose to belong to. This 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 untrustworthy looks, 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. In repeated public goods games, people have the 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, 2004b). 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 these 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 almost disappear 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.

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 by combining identification and information on past actions cooperation in public goods games increases (Andreoni and Petrie, 2004b). 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, 2004a; Petrie, 2004; and Eckel and Wilson, 2003), and attractive people make more money (Mobius and Rosenblat, 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 distinguished 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 is relies only on the comparative static properties of the equilibrium with discrimination.

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. Observable characteristics of participants are important insofar as they reveal information on likely strategies to be played. In equilibrium, people will play 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 reflects 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. 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.

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 Lima is full of internet cafes. As a result, there is a high proportion of the population with expertise using the internet. According to a survey conducted in 2003, Lima had 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. Students belong to a potentially very unique segment of the population, thus reducing the external validity of the results and impeding us from drawing clear policy implications.

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 were representative of the young working population in metropolitan Lima. With these objectives in mind we implement a four step 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, 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 homogenously across gender and income levels.

Secondly, based on these criteria we contracted two specialized companies 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 data bases. This 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 comply with all of our criteria. From the resulting sub-sample we performed a random lottery and select the individuals to be part of the experiment.

Thirdly, and with 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 invite subjects. Second, it is one of the biggest SME clusters 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% of the population of La Victoria (where Gamarra is located) was 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 identify the blocks with the highest concentration of commercial buildings. This is 3 long blocks (approximately 8 standard 100mt blocks) and crossed by three transversal blocks. 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.

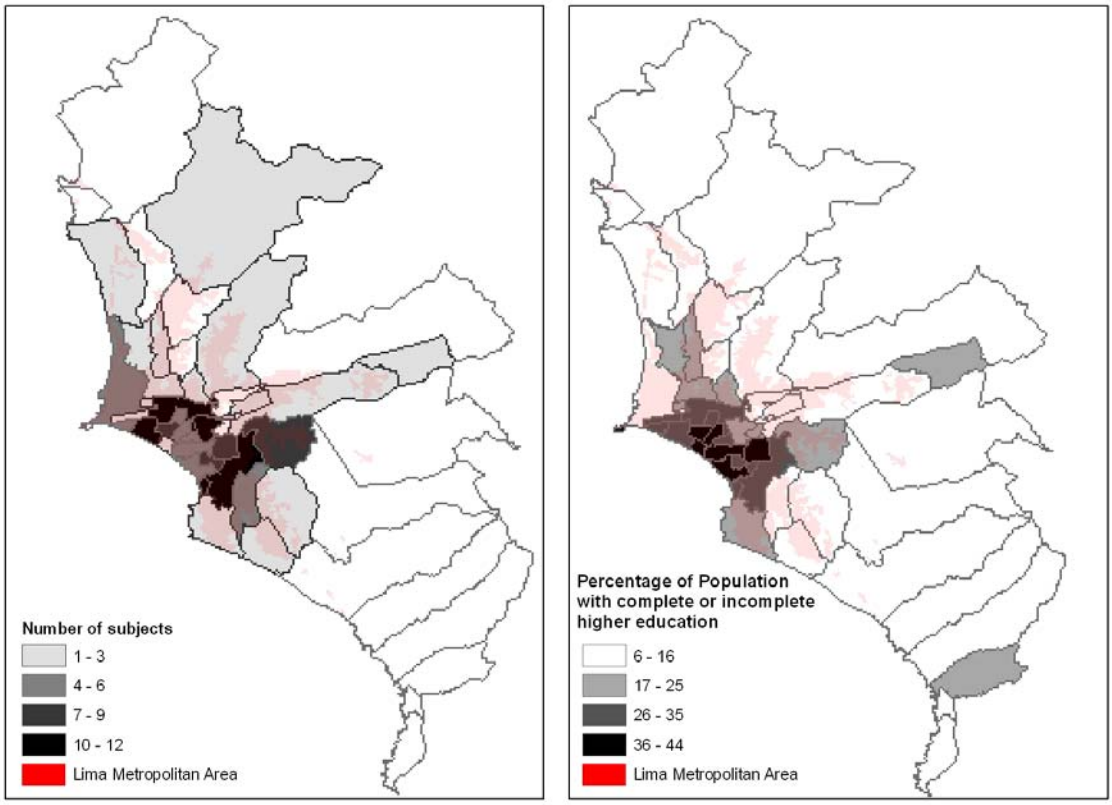
Finally, to be representative of subjects of university age and that are part of the university system, we also invite potential subjects that were taking classes at extension institutes within the major universities located in metropolitan Lima. These universities have student bodies that represent 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 already are 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 if presented in a graphical manner. The experiments in this research required simply that the subjects know how to use a computer mouse.

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 built from the *Encuesta Nacional de Hogares* (ENAHO) 2004. This is a sub-sample that 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 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. Our sample is slightly more educated and had significant 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. However, 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 [1/.]
Age	26.31	27.95
Male	61.6%	54.6%
Level of education [2/.]		
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 [2/.]	14.61	14.02
Native language [3/.]		
Quechua	0.6%	1.5%
Aymara		
Spanish	95.0%	98.5%
Other	4.4%	
Migrant [4/.]	15.2%	37.9%
Weekly hours of work [5/.]	35.4	44.7
Monthly income [5/. 6/.]	1087.3	1099.2
Health insurance [7/.]	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 in average, and most of them had certain type of higher education. The average number of years of schooling was 14.6, with a standard deviation of 1.9. Close to 70% of the sample had complete or incomplete university studies. Only 20% of the individuals graduated from a public school while 45% from a private religious high-school. Their parents' education was on average lower than their own. Even though, 45% of the fathers and 27% of the mothers finished university studies.

All the applicants had some labor experience, 98% worked during the last twelve 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 done some technical or professional studies at a university and whose parents had achieved post-high-school diplomas, either at universities or vocational institutions.

With regard to asset ownership, we found that most of them live dwellings with more than two rooms, and 40% of the sample lives in dwellings of 5 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 they belong to different socioeconomic levels.

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 centimos to the subject.[1] 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 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, 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 would they prefer to be in their group, we create the following procedure. Subjects rank all other subjects from most preferred to 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

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

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 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. 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 than 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 is additive, it does not matter whether he is chosen first or last.

Additivity of preferences over others' company may be a strong assumption. Some combination of people might be less successful than others. For instance, some women might be very cooperative with other women but not so with men. 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 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 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. 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 Pacific University 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).

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 at subjects in the experiment, to rate the photos of the subjects in terms of race, beauty and also height. A rater only rated the photo in terms of one characteristic, race, beauty or height, 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 (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 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 ten men and ten 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.[2]

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 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, beauty or height to get the final ratings we use to analyze the data.

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% 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 of these two dimensions in our analysis of contributions and ranking in the next section.

[2] 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.

Results & Discussion

Table 2 present descriptive statistics of the experimental subjects.[3] Three out of 5 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 3 more years of post-secondary education and 29% have a college degree. The sample also reflects the different ethnic and cultural background embedded in the population of Lima. 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 5 persons.

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[1]	0.17	0.60	0	4
Indigenous Grandparents (n)	157[1]	0.31	0.89	0	4
Height (meters)	160	1.69	0.08	1.52	1.94
Family Size	156[1]	4.95	2.24	1	13
1[Religious High School]	160	0.44	0.50	0	1

[1] There are three missing post-experiment questionnaires from the sample.

What Did People Do in the Experiment?

Figure 1 and Figure 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% of subjects' endowment for low types in the Two Types treatment to 46% 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% for low types in the Two Types treatment and to 27% 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 in Contribution Only to 46% for low types in Two Types. Contributions, in the last round of the second sequence, decrease to 14% in Photo Only and to 32% for low types in Two Types.

[3] There are 3 post-experiment surveys missing from the sample

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 graph also shows convergence towards play of dominant strategies by high types.

Figure 1

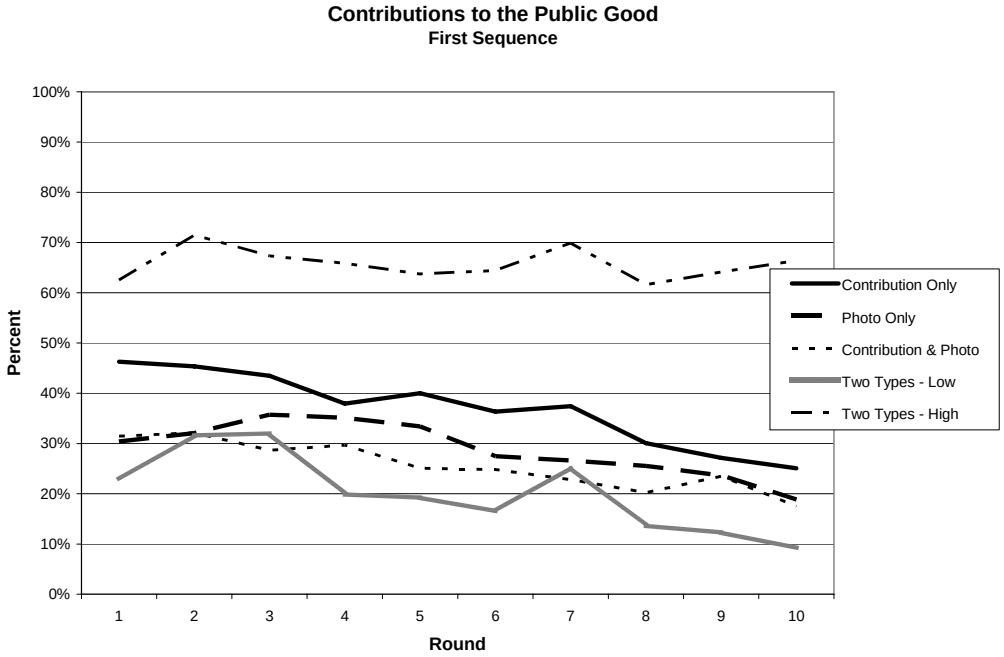
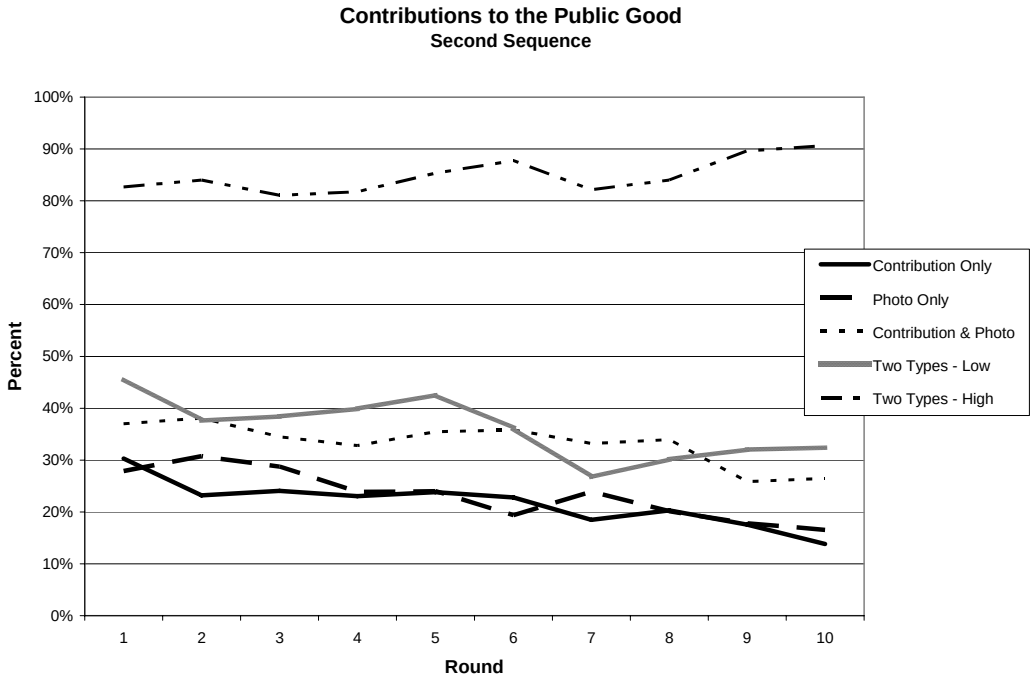


Figure 2



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 rough market conditions, making them behave more selfishly. Or, 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.[4]

Table 3. Percent of Endowment Contributed to the Public Good (sequence 2)			
Variables	Contribution Only, Photo Only, & Contribution and Photo	Two Types	All Treatments
1[Men]	0.05 (0.22)	-0.03 (0.75)	0.03 (0.38)
Age (years)	0.00 (0.71)	-0.01 (0.25)	-0.00 (0.81)
Education (years)	0.00 (0.58)	-0.03 (0.13)	-0.01 (0.47)
Height (meters)	-0.03 (0.90)	1.15 (0.04)	0.26 (0.25)
White Intensity	-0.024 (0.37)	-0.00 (0.97)	-0.03 (0.32)
Indigenous Intensity	-0.02 (0.52)	-0.00 (0.94)	-0.02 (0.49)
1[Religious High School]	-0.01 (0.67)	0.03 (0.67)	0.01 (0.69)
1[Low Type]	-	-	0.27 (0.01)
1[High Type]	-	0.47 (0.00)	0.71 (0.00)
Round	-0.01 (0.00)	-0.00 (0.73)	-0.01 (0.00)
Constant	0.27 (0.50)	-0.73 (0.40)	0.03 (0.94)
Individual Random Effects	*Yes*	*Yes*	*Yes*
Group Fixed Effects	*Yes*	*Yes*	*Yes*
R2			
N	1200	400	1600
p-values in parentheses			

The regressions in Table 3 show that behavior is not correlated with personal characteristics. On average, contributions decrease by 10% from round 1 to round 10. There is slight effect of taller people giving more in the Two Types treatment. It is 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.

[4] Results are robust to different specifications.

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.

How Were People Ranked?

The previous section shows that there is little evidence supporting 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 describe in the section on Race Classification. We use this aggregated racial intensities to create a discrete variable determining if a person is white or indigenous. A person is considered white if her average racial intensity in the white dimension is above the median and her average racial intensity 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 for each treatment separately.[5] 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 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. The race variables are discrete variables that indicate if the person falls in the top half of the distribution in one racial dimension and the bottom half in the other.

[5] 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.

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% 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. Ten 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 that height itself explain 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 investigates the presence of discrimination across treatments.[6] Table 7 shows a linear probability model on the likelihood to be in the top 4 of any list. As mentioned above being a man decreases the probability of being among the top 4 and height increases the probability of being among the top 4. Table 8 shows the likelihood of being in the bottom 4 of any list. Both of these tables confirm previous results.

Table 7. Probability of Making it to 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 Making it to 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 4. Also, looking indigenous increases the likelihood of being names among the bottom 4.

[6] The results in Tables 6 through 8 are similar if we use probit or robust standard errors.

Conclusions & 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, be it ethnicity or 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 vanish almost completely once information on others' behavior is provided.

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. This evidence cannot be reconciled with stereotyping and therefore indicates that some level of preference-based discrimination is present in Peru. Intriguingly, while tall women are preferred in the absence of information, they are less likely to making to 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 agreements 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 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 give precise and reliable measure of one's performance. That is, it is not clear that policy makers have the tools to make signals clearer or to make measurement of performance in the work place better. It is entirely possible also that while discrimination in the work place is diminished through public intervention other avenues survive like marriage or neighborhood sorting.

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

References

Andreoni, James and Ragan Petrie, "Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments," Working Paper, Georgia State University, January 2004.

Andreoni, James and Ragan Petrie, "Public Goods Experiments Without Confidentiality: A Glimpse Into Fund-Raising," *Journal of Public Economics*, 2004, 88(7-8), 1605-1623.

Bohnet, Iris and Bruno Frey, "The Sound of Silence in Prisoner's Dilemma and Dictator Games," *Journal of Economic Behavior and Organization*, 1999, 38, 43-57.

Burnham, Terence C., "Engineering Altruism: a Theoretical and Experimental Investigation of Anonymity and Gift Giving," *Journal of Economic Behavior and Organization*, 2003, 50, 133-144.

Carter, Michael and Marco Castillo, "An Experimental Approach to Social Capital in South Africa," Working Paper, University of Wisconsin-Madison, 2003.

Castillo, Marco and Ragan Petrie, "Discrimination in the Warplace: Evidence from a Civil War in Peru," Working Paper, Georgia Institute of Technology, 2005.

Castillo, Marco and Ragan Petrie, "Discrimination in the Lab: Experiments Exploring the Impact of Performance and Appearance on Sorting," Working Paper, Georgia State University, 2005.

Cummings, Ronald and Paul Ferraro, "Inter-cultural Discrimination in the Ultimatum Game: Ethnic Bias and Statistical Discrimination," Working Paper, Georgia State University, 2004.

Eckel, Catherine and Rick Wilson, "Conditional Trust: Sex, Race and Facial Expressions in a Trust Game," Working Paper, Rice University, 2003.

Kagel, John and Alvin Roth, *Handbook of Experimental Economics*, 1995.

Hammermesh, Daniel and Jeff Biddle, "Beauty and the Labor Market," *American Economic Review*, 1994, 84(5), 1174-94.

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

Manski, Charles, "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, Vol. 60, No. 3. (Jul., 1993), pp. 531-542.

Mobius, Markus and Tanya Rosenblatt, "Why Beauty Matters," forthcoming *American Economic Review*, 2005.

Ñopo, Hugo, Saavedra, Jaime, and Torero, Maximo; (2005). "Ethnicity and Earnings in Urban Peru," re-submitted to *Economic Development and Cultural Change* and previously IZA DP No. 980, 2003.

Petrie, Ragan, "Trusting Appearances and Reciprocating Looks: Experimental Evidence on Gender and Race Preferences," Working Paper, Georgia State University, February 2004.

Riach, Peter A. and Judy Rich, "Field Experiments of Discrimination in the Market Place," *Economic Journal,* CXII (2002), F480–F518.

Torero, Maximol Jaime Saavedra, Hugo Nopo and Javier 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.