Beauty, Gender and Stereotypes: Evidence from Laboratory Experiments

James Andreoni
University of Wisconsin
Madison, WI 53706

Ragan Petrie*
Georgia State University
Atlanta, GA 30303

October 2005

Abstract

The existence of a beauty premium in the labor market and the male-female wage gap suggests that appearance can matter in the real world. We explore beauty and gender in a public goods experiment and find similar effects. We find a beauty premium, even though beautiful people contribute, on average, no more or less than others. The beauty premium, however, disappears when we provide information on individual contributions, and becomes a beauty penalty. Players seem to expect beautiful people to be more cooperative. Relative to these expectations, they appear more selfish, which in turn results in less cooperation by others. These appear to be clear examples of stereotyping. We also find a substantial benefit to being male, especially with information. This is primarily due to men being better "leaders." Men tend to make large contributions, and people follow their example and give more in later rounds.

1. Introduction

It is well known from labor market studies that beauty and gender can have big effects on earnings. Hamermesh and Biddle (1994) have shown a significant premium to beauty, with attractive people earning more money than unattractive people. There is also a significant and persistent male-female wage gap. Even when controlling for age and experience, men earn about 25% more than women (O'Neill, 2003, 1998). While some of these differences can be attributed to labor market factors, much of the beauty premium and wage gap remains unexplained.

With labor market studies as inspiration, we look at the returns to beauty and gender in an economic laboratory experiment where there are benefits to group cooperation. Typically economic experiments take great pains to shield the identities of subjects from each other, and as such have nothing to say about how appearances may affect earnings.[1] Instead, we reveal the identities of players to one another by showing their digital photos in the experiment. We find that beauty and gender have significant and sometimes unexpected affects on earnings. Furthermore, the behaviors and apparent stereotyping we find may provide some clues into why gender and beauty are so important in the labor market.

The experimental setting we consider is a repeated linear public goods game. While not a direct test of the beauty premium or the wage-gap found in the labor market, a public goods game is nonetheless an interesting institution for exploring how such wage differences can emerge in an employment setting. People

---

[1] An important exception to this is research by Mobius and Rosenblat (2003) who use a labor market experiment to decompose the beauty premium. In their experiment, an employer receives a precise measure of a worker's productivity from a practice round and decides a wage to pay the worker. The interaction between the employer and worker is one shot, so the employer cannot verify if a beautiful worker was as productive as originally thought.

often work in teams where shirking cannot always be perfectly monitored. This allows for stereotyping to color evaluations of and reactions to both free riding and generosity, and to affect the productivity of the team. A repeated public goods game offers a setting where stereotyping is possible (in groups of more than two) and people have an opportunity to see if their expectations of behavior mesh with actual behavior. When effort is observable, certain people can "set the tone" for the work group. Gender and beauty may affect which people in the group may be emulated by others, how much retribution people take against shirkers and how much cooperation is used to reward generosity. In the end, the beauty and gender of the individuals in the group will affect the cooperation and success of the group as a whole, and the earnings of its individual members.

In our experiment we show each player the digital photos of all other members of their group each round. We have two conditions, one in which only total group contributions are revealed, and another in which information on each player's contribution is revealed.

Our experiments find evidence for a beauty premium. This premium, however, disappears once people know exactly what each group member contributed to the public good. When only the total group contribution is observable, attractive men and women make more money than unattractive men and women, even though they are no more or less cooperative, on average, than unattractive people. When individual contributions are observable, the reward to being beautiful disappears. People seem to expect beautiful people to be more cooperative than others, and when their behavior does not meet expectations, people are less cooperative with them. There is also a difference in payoffs for men and women, but not always favoring men. Women make more money than men when only group contributions

2

are known. This can be attributed primarily to the stereotype that women are more helpful. Men, however, do best when individual contributions to the group are clearly identified. Men earn 15% more when individual contributions are known, compared to when only group contributions are known. Interestingly, women's payoffs do not change. This effect can be attributed entirely to men being better "leaders." They contribute their full endowment more often, and others follow their good example by contributing more in later rounds.

With these results, we begin to see some of the underlying factors that could generate both a beauty premium and a wage gap, especially in team work environments. People give beautiful people the benefit of the doubt in groups and cooperate more with them, thereby enhancing group welfare. Women earn more than men when individual contributions to group output are unknown because they are stereotyped overall to be more helpful. However, once individual contributions are known with certainty, the beauty premium disappears and a male premium replaces it. This appears to be due to people being more willing to follow the lead of generous men. When a man's generous contribution is observed, it engenders more cooperation than a similar contribution by a woman. This effect is amplified when the proportion of men in the group is higher, suggesting a kind of generosity-competition among men. As a result, men earn more money in the experiment when individual contributions to group output are known. In our data, the male premium is almost equal in size to the beauty premium.

Why are these results important? They illustrate how easily subtle sex and beauty stereotypes can influence economic outcomes, and suggests that further study on the role of stereotypes in economics could yield important insights.

2. Background

In our experiment subjects see the digital photos of their partners on the computer screen when they make their decisions. In half of our sessions we also tell subjects what each member contributed in the prior round. There are three ways this could have an effect. Simply showing faces could matter, gender and beauty could matter, and finally the information on amounts given could matter.

There is good reason to expect that simply seeing the faces of partners will affect play. Bohnet and Frey (1999a,b) find that visual identification increases cooperation in one-shot Prisoner's Dilemma and Dictator games, and Burnham (2003) reports that giving more than doubles in Dictator games when at least one of the partners sees the other's photograph. Scharlemann, Eckel, Kacelnik, and Wilson (2001) see slightly more cooperation when subjects are faced with a photograph of smiling bargaining partners, and Eckel and Wilson (2003) find subjects are more trusting when they see their partner. Solnick and Schweitzer (1999) report a significant effect of attractiveness on offers received and responses made in ultimatum bargaining games.

Gender also has been shown to have an important effect on behavior and payoffs. Eckel and Grossman (1998) find groups of women to be more generous than groups of men in Dictator games, while Andreoni and Vesterlund (2001) find women more equalitarian than men in Dictator Games. Buchan, Croson, and Solnick (2003) report that women are less trusting but more trustworthy than men in an investment game. Gneezy, Niederle, and Rustichini (2003) find women to be more productive in competitive-pay environments when in same-sex groups,

rather than mixed-sex groups.[2]

There is also experimental evidence from the sociology and psychology literature that beauty carries a premium. In Prisoner's Dilemma games, people are more cooperative with attractive partners (Mulford, Orbell, Shatto, and Stockard, 1998). Indeed, in a comprehensive review of the literature on beauty, Langlois, Klakanis, Rubenstein, Larson, Hallam and Smoot (2000) find that attractive people are not only judged and treated more favorably but they also behave differently.

Finally, simply knowing what each other player has chosen can also have an impact. For instance, Andreoni and Petrie (2004) find that cooperation is highest when actions can be linked to the actor. But, information may also help adjust judgements. Social expectancy theory predicts that if people have different expectations of how, for instance, attractive and unattractive others will behave, then they may have totally different reactions to the same observed behavior depending on which person they are facing (Darley and Fazio, 1980; Zebrowitz, 1997). Furthermore, psychologists suggest that people come to the table with prejudices and stereotypes. However, after meeting someone and gaining more information, these prejudices are often revised or washed away (Devine, 1989). Related to this are issues of similarity and familiarity. People tend to favor those similar to themselves (McPherson, Smith-Lovin and Cook, 2001), and working with familiar others can reduce transaction costs, as familiarity can enhance trust (Glaeser, Laibson, Scheinkman, and Soutter, 2000).

This suggests that people may behave differently with others when they know their decisions, rather than just their appearance, and that people may want to emulate the choices of "similar" others. Also, people can use their choices to signal

---

[2]See Eckel and Grossman (2000) for a review of gender differences in laboratory experiments.

5

information to others in order to establish a link later (such as getting a date). Identification may not only serve to help form impressions before actions, but people may also use identification as a guideline for judging actions and formulating a response.

3. Methodology

Our experiments use a linear public goods game. The game is repeated, and subjects make their decisions on a computer. Each session has 20 subjects, and they are randomly divided into 4 groups of 5 subjects. In each round a subject is endowed with 20 tokens that could be invested in a private good or a public good. The private good pays $0.02 per token invested by the individual, and the public good pays $0.01 per token invested by the entire group. Therefore, the marginal return to investing in the public good is 0.5. Each subject is paid based on his investment in the private good and the total group investment in the public good. Subjects play with the same group of 5 people for 8 rounds, then are randomly re-matched to new groups for another 8 rounds, until they play with five different groups, each for 8 rounds. So, subjects play 40 rounds in total in each session.[3]

A digital photograph is taken of each subject at the beginning of the experimental session. At the end of each round, the total amount contributed to the public good by all group members is reported. There are two treatments: No Information and Information.[4] In the Information treatment, both the photo and the contribution of that group member from the previous round are displayed.

---

[3]Very few subjects knew other subjects in the experiment (87% of the subjects had "never met" any other subject in the experiment).

[4]Note that, in Andreoni and Petrie (2004), the No Information treatment is called "Photos" and the Information treatment is called "Information-and-Photos."

The contribution is listed below each photo, and the photos are rearranged each round with the highest contributor on the left and the lowest on the right. In the No Information treatment, only the photos of each group member are displayed on the screen and no information on individual actions is given.

All treatments were conducted twice. This gives us 80 subjects in total. Subjects were recruited from economics and business classes at the University of Wisconsin - Madison. There were 39 female subjects and 41 male subjects in total, ranging in age from 18-29 years (mean 20.3 years old). Each experimental session took about an hour and a half, and the average subject payment was $26.46 (standard deviation $3.13).

Subsequently, people not involved in the experiments rated each photo as to the person's physical attractiveness or how helpful looking he/she appears. In total, the raters saw 140 photographs. These photos were of the 80 subjects in the No Information and Information treatments and 60 other photos of subjects in two other public goods game treatments.[5] These other treatments were not used here as they do not allow a clear test of beauty and gender effects.

We chose to have the raters view the 80 subjects randomly mixed among the other 60 photos to get a "global" measure of physical attractiveness and helpfulness. We want to know if a subject is considered physically attractive among his or her peers, not only among the subjects in the session. Drawing on this broader distribution of 140 photographs gives us more information and confidence in our classification and subsequent results.

---

[5] The photos of the 60 other subjects come from the same cohort as the 80 subjects in the No Information and Information treatments. They were recruited at the same time as the 80 other subjects from economics and business school classes at the University of Wisconsin-Madison. Of the 60 subjects, 33 are male and 27 are female, ranging in age from 18-28 years (mean: 20.1 years old).

Each rater rates all 140 photographs along one dimension, physical attractiveness or helpfulness. The raters view the photos on the computer in a random order, assign a rating, and have the ability to move back to previously rated photos to change or check ratings. Raters know they can do this before they begin rating. The raters use a 9-point scale, with 1 indicating "not at all" and 9 indicating "very much so." Raters are told to think of the number 5 as average. There are 15 raters in total. Four women and four men rated each photo in terms of physical attractiveness, and four women and three men rated each photo in terms of helpfulness. The raters were all undergraduates of the University of Wisconsin-Madison, with an average age of 20.7. The raters were paid a flat fee of $15 to rate 140 photographs, and the rating task took 30-45 minutes to complete.

4. Beauty Classification

In this section we discuss our beauty classification methodology, the reliability across those rating beauty, and explain how we define attractiveness.

First, we assume the raters use all 140 photos as a frame of reference. When discussing consensus among raters, therefore, we consider the consensus for all 140 photos. Also, when defining attractiveness, we use all 140 photos as the entire frame of reference.[6]

Turning to the consensus among raters, the inter-rater reliability for physical attractiveness is 0.86, and for helpfulness it is 0.82.[7] These reliability measures

---

[6] Note that similar, yet not as strong, results hold if we define the frame of reference to be only the 80 photos of the subjects in the Information and No Information treatments. So, it appears that the frame of reference is important in defining who is attractive. We opt to base the analysis of this paper on the larger frame of reference because it improves the accuracy and predictive power of our ratings.

[7] We use the Cronbach coefficient alpha for standardized variables (Cortina, 1993; SAS Insti-

compare well with previous research on attractiveness and suggest a strong consensus among raters. Our reliability measures also compare well with previous research on attractiveness, where the range of reliability measures is between 0.85 and 0.95 (Langlois et al., 2000).[8]

There is some variability in the spread and average rating any given rater gave. Some raters use the entire range from 1 to 9, but others truncate the range between 3 and 8. Therefore, each rater's ratings of all photos are standardized by the rater's mean and standard deviation to mean zero and standard deviation of one. Then, for each subject, we average the standardized ratings across all raters to give a value of physical attractiveness and helpfulness for each subject. These averaged standardized ratings are used throughout the paper.

Overall, women are rated significantly higher in attractiveness than men. Women receive an average standardized rating of 0.29 and men -0.26. The same holds for helpfulness. Women were rated, on average, at 0.25 and men at -0.22. That is, women are considered more attractive and more helpful-looking than men.

There are, obviously, many ways of defining an attractive person. Because we wish to examine the effects of beauty on earnings and women are disproportionately rated more attractive than men, we take into consideration both gender and beauty rating in our classification. Of the 140 subjects, 66 are women and 74 are

tute Inc., 1988). This measures the correlation between all raters and adjusts for rater specific mean and variance. While this rating raises automatically as the number of raters increases (we thank Daniel Hamermesh for pointing this out), we report it in order to compare with other studies using attractiveness ratings. Using an alternative measure of reliability, an intra-class correlation coefficent (two-way random effects model), we find similar results. The ICC coefficient is 0.7506.

[8]Langlois et al. (2000) report an acceptable range of reliability measures is between 0.85 and 0.95. The fact that measures fall on the lower end of this range and reflect our 9–point scale. Most scales were 5-point or 7-point.

men. For women, we took the top quartile of the distribution of attractiveness ratings for these 66 women and classified them as attractive. We did the same for the men. Then, we took the bottom quartile of the distribution of attractiveness ratings for the 66 women and classified them as unattractive. We did the same for men. Anyone in the middle of the distribution was classified as middle attractive. For robustness, we repeated our analysis with both more and less exclusive definitions, with nearly identical results.[9] This classification of subjects is also within the range used by Hamermesh and Biddle (1994), where between 25%-34% of their samples were classified as above average or higher in attractiveness.[10]

Table 1
Distribution of Attractiveness

	Women	Men
Attractive	8	8
Middle	19	25
Unattractive	12	8
Total	39	41

The distribution for the 80 subjects considered here is shown in Table 1.[11] By definition, there are roughly an equal number of men and women classified as attractive.[12]

---

[9]We also classified subjects as attractive and unattractive by using the 20th-percentile and the 30th-percentile. Results using the 20th-percentile cut-off are strikingly similar to those using the quartile cut-off. Results using the 30th-percentile cut-off follow similar trends to those using the quartile cut-off, however some results that are statistically significant with the quartile cut-off are not with the 30th-percentile cut-off.

[10]Note that Hamermesh and Biddle (1994) did not use the same range of their data to classify people as unattractive. Between 7%-17% of their samples were classified as below average or lower in attractiveness.

[11]When we did not condition on gender, the distribution of attractive, middle attractive and unattractive subjects across men and women was skewed. 14 women and only 3 men were classified as attractive, and 14 men and only 4 women were classified as unattractive.

[12]The distributions of attractiveness are similar across the 140 subjects and the sample of 80 subjects. The hypothesis of independence across the two distributions cannot be rejected by a Chi-Square test (p-value=0.761).

5. Results

We first discuss the beauty premium, then gender differences in payoffs, and finally the interaction of beauty and gender.

5.1. Does Beauty Pay?

First consider the No Information treatment. The left-hand side of Table 2 shows that there are positive returns to beauty. Attractive people make 9% more than middle attractive people and 15% more than unattractive people. The returns to beauty are similar to those found by Hamermesh and Biddle (1994), where above-average attractive people earned about 5% more than average-attractive people. The differences across the three attractiveness categories are significant.[13]

Table 2
Average Payoff

	No Information	**Information**
Attractive	$27.49	$27.57
Middle	$25.25	$27.37
Unattractive	$23.88	$27.77
p-value[*]	0.0425	0.7865

*Joint Kruskal-Wallis rank sum test for equality of payoffs across the three categories.

While there is evidence for a beauty premium in the No Information treatment, it disappears in the Information treatment. The right-hand side of Table 2 illustrates this. In the Information treatment, payoffs appear independent of beauty.[14]

---

[13]In pair-wise tests using the Wilcoxon rank sum test, attractive people make significantly more than middle attractive (p-value=0.0905) and unattractive people (p-value=0.0108). However, middle attractive people do not make significantly more than unattractive people.

[14]In pair-wise comparisons using a Wilcoxon rank sum test, payoffs are not significantly different between attractive, middle attractive, or unattractive people.

We see that returns to beauty are significant, but only when performance cannot be observed. Both attractive men and women earn more money. We explore the returns to gender further in section 5.2. Next, we ask why the beauty premium might exist.

5.1.1. Behavior of Attractive People

Why do attractive men and women make more money? Two possible explanations are that they are either more selfish or that they engender more cooperation from others. The first explanation is explored in Table 3. For the No Information treatment, the left-hand side of the table shows that attractive people and unattractive people are, on average, less selfish than middle attractive people. Both attractive and unattractive people contribute between 47-55% of their endowment to the public good, but middle attractive people only contribute 31%. While the average contribution of an attractive person is significantly different than that of a middle attractive person, there is no significant difference in contributions of attractive and unattractive people.[15]

If higher earnings are due to more selfishness, then it makes sense that middle attractive people make more money than unattractive people since they contribute less on average. This also would imply that attractive people would make *less* money than middle attractive people. But, attractive people make more. While there is no significant difference between the average contribution of an attractive and unattractive person, an attractive person makes $3.61 more than an unattractive person. Selfish behavior is *not* an explanation for why attractive people make more money, but it may explain the higher payoff for middle attractive people

---

[15]In pair-wise comparisons using a Wilcoxon rank sum test, the difference in the average contribution of an attractive person and an unattractive person is insignificant (p-value=0.5966).

relative to unattractive people.

Table 3
Average Percent Contributed per Round

	No Information	**Information**
Attractive	55.0	42.1
Middle	31.2	49.3
Unattractive	46.9	50.0
p-value∗	0.0084	0.5872

∗Joint Kruskal-Wallis rank sum test for equality of contributions
across the three categories.

For the Information treatment, the right-hand side of Table 3 shows that attractive people contribute less than middle attractive and unattractive people. However, the difference in contributions is not significant, nor are their payoffs in Table 2.[16]

We do not find evidence that the beauty premium observed in the No Information treatment is due to more selfish behavior by attractive people. Indeed, there are no clear trends in contribution behavior and payoffs. Attractive people are no more cooperative than unattractive people, but they still make more money. This suggest that the difference in payoffs may be due to how others respond to beauty. We explore this next.

5.1.2. Favoritism and Beauty Stereotypes

Could the observed beauty premium be due to favoritism towards attractive people? If favoritism exists, then we would expect people to react more cooperatively with attractive people than with people who are not attractive. Table 4 explores this hypothesis. We compare the average contribution of nonattractive people

---

[16]In pair-wise comparisons using a Wilcoxon rank sum test, there are no significant differences in the average contribution of attractive, middle attractive, or unattractive people.

(those who are classified as middle and unattractive) in groups with no attractive people and in groups with at least one attractive person. We concentrate on the contributions of nonattractive people because we want to isolate the effects of beauty in a group on the behavior of others.

Table 4
Average Percent Contributed by Nonattractive People
in groups of Attractive People, All Rounds

# of Attractive people in group	No Information	Information	All
0	29.8	53.8	42.2
1 or more[*]	41.4	44.8	43.0
p-value	0.0023	0.0234	0.7477

*In the No Information treatment, there are 14 groups with one attractive person, 6 with two attractive people, 2 with three attractive people, and 2 with four attractive people. In the Information condition, there are 11 groups with one attractive person, 7 with two attractive people, and 5 with three attractive people.

Table 4 shows that the contributions of nonattractive people depend on the beauty of the other group members. In the No Information treatment, the presence of beauty in the group does engender more cooperation. The average percent contributed in groups with no attractive people is 29.8%, but in groups with at least one attractive person, the average increases to 41.4%.[17] This difference is significant and suggests the presence of beauty may engender more cooperation. In the Information treatment, however, it is the absence of beauty that engenders more cooperation. The average contribution in groups with at least one attractive person is 44.8%, but in groups with no attractive people, it increases to 53.8%. This difference is also significant, but it contradicts favoritism.

---

[17]Similar results hold if we include the contributions of attractive people. In this case, contributions increase to 46.2% in the No Information treatment and decrease to 44.8% in the Information treatment. This suggests that it is not merely nonattractive people who are affected by the presence of attactive people in the group.

14

Why do nonattractive people reverse their behavior from one treatment to the next? One hypothesis is rooted in beauty stereotypes and reciprocity. That is, people are more cooperative with attractive people because they believe attractive people are helpful. Langlois et al. (2000) found, in a comprehensive review of the beauty literature, that attractive people are consistently judged and treated more positively. We also find this in our ratings data. Thirty-eight percent of attractive subjects are also judged helpful looking, 18% of middle attractive people are also judged helpful looking, but only 5% of unattractive people are judged helpful looking. This is remarkable because these ratings are by people who did *not* observe the decisions of the individuals in the photos *and* they did not rate the subjects in terms of attractiveness. This suggests that people independently *expect* attractive people to be more cooperative.

While there may be reason to believe such a stereotype exists, it is a bit tricky to tease apart because people may hold stereotypes but not act upon them. However, Devine (1989) argues that if people hold stereotypes, once they meet a stereotyped individual and gain more information, they revise or eliminate the stereotype. That is, without information, they act upon the stereotype, but with information, behavior changes. How behavior changes depends on how people revise or react to behavior that does not coincide with the stereotype.

If people hold the stereotype that attractive people are cooperative and people wish to reciprocate cooperation, then we should see changes in behavior in two ways. First, without information, people should be more cooperative with attractive people. Evidence for this behavioral change can be found in Table 4. Without information, we see that nonattractive people are more cooperative in groups with attractive people than in groups with no attractive people, even though we know

15

that attractive people contribute the same, on average, as unattractive people. Moreover, there is higher cooperation in groups with attractive people in every round of the 8-round sequence, with the difference increasing over the rounds. In round one, the difference in cooperation is 10.4 percentage points, and in round eight, it is 18.0 percentage points.

Second, with information, we should see differences in behavior in groups with attractive people and in groups without attractive people because people can now compare how the actions of attractive people differ from their expectations. How behavior changes depends on whether deviations from expectations are rewarded or punished.[18] In Table 4, we see that with information, nonattractive people are more cooperative in groups with nonattractive people. They contribute 20% more. This result suggests stereotyping and punishment of behavior that does not meet expectations since there was no significant difference in contribution levels of attractive, middle attractive, and unattractive people in the Information treatment.

What is remarkable about this result is that if people base reciprocity solely on met or unmet expectations on behavior, then there should be no difference in Table 4. The presence of an attractive person in the group should not matter. That it does suggests that people react differently to the same actions of attractive and unattractive people.[19]

---

[18]Bosman and van Winden (2002) find that subjects that expected their partners to share more of the pie than they actually did were more likely to punish in a power-to-take game.

[19]In the very first round of the experiment, before any experience is gained with attractive people, there is no significant difference across treatments in the average contribution of nonattractive people in groups with attractive people. This suggests that subjects start out with the same beliefs over the actions of attractive people.

In sum, it appears that people discriminate based on beauty stereotypes. Attractive men and women are given the benefit of the doubt when group members do not see performance, even though they contribute amounts similar to unattractive men and women. Indeed, the presence of beauty engenders more mutual cooperation, and this may explain the beauty premium. However, when group members can evaluate performance, the presence of beauty engenders less cooperation. People seem to expect attractive men and women to be more cooperative than average. When they are not, others are less cooperative with them, and the beauty premium disappears.

5.1.3. Gender Differences in Beauty Stereotypes

Does favoritism differ by gender? In the No Information treatment, it does not. Both nonattractive men and women are significantly more cooperative in groups with attractive people than in groups with no attractive people.

In the Information treatment, only men significantly decrease their contributions in groups of attractive people compared to groups with no attractive people. In groups with no attractive people, men contribute 59.9%, but this declines to 41.1% in groups with at least one attractive person. Women increase their contributions, from 46.3% to 47.7%, but this is not significant.[20] This may mean that nonattractive men and women may hold different beauty stereotypes. Or, they may hold the same stereotype, but only men act upon it.

These results suggest that there are no gender differences in favoritism, but there may be in beauty stereotypes. In the No Information treatment, both men and women are more cooperative with an attractive person in the group. Men

---

[20]Strikingly similar results hold if attractive men and women are included in the average.

17

and women, however, react differently to information and attractive people in the group. In the Information treatment, the reduction in contributions in groups of attractive people is driven by the actions of men.[21]

5.2. Does Gender Pay?

We now move the focus from beauty to gender. Overall, men and women make the same amount of money. Table 5 shows that both men and women make around $26.50 over both treatments. However, women make $1.67 more than men in the No Information treatment, and men make $1.38 more than women in the Information treatment. The former difference is significant at the 10% level, and the latter is not.

Table 5
Average Payoff

	Women	**Men**	**p-value**[*]
Both Treatments	$26.50	$26.41	0.9463
No Information	$26.22	$24.55	0.0962
Information	$26.80	$28.18	0.1162
p-value[*]	0.5273	0.0002	

*Pair-wise Wilcoxon test for equality of payoffs across the two
   treatments/genders.

Within gender, women make about the same across treatments. They make $26.22 in No Information and $26.80 in Information. By contrast, the payoff for men is 15% higher in the Information treatment compared to the No Information treatment. This difference is significant.

---

[21]Fershtman and Gneezy (2001) also find that men are more apt to hold and act on stereotypes. Using trust experiments, they find that men in Isreali society are more likely to act on ethnic stereotypes than women.

There is a gender difference, but that switches from favoring women to men when actions are known. There are large benefits to information, but only men reap them. Again, there are two possible explanations. Women could be more selfish than men, or men may engender more cooperation with their performance than women. We discuss these possibilities next.

5.2.1. Behavior of Men and Women

Are women more selfish than men? Table 6 shows that, on average, women are no more or less cooperative than men—men give roughly 15% more than women, but this is not significant. However, men play the extremes significantly more than women.[22] Men contribute zero tokens 27.7% of the time, but women do so only 16.3% of the time. Women also contribute all of their tokens only 9.4% of the time, compared to men who do so 22.1% of the time. Similar results hold for each treatment separately.

Table 6
Average Contribution, Zero Tokens, and All Tokens
Both Treatments

	Men	Women	P-value
Average Contribution (%)	46.7	40.7	0.1938
% of Times Zero Tokens	27.7	16.3	0.0594
% of Times All Tokens	22.1	9.4	0.0012

This suggests that the gender premium to women is not due solely to more selfishness on their part. Women seem to be benefiting from a combination of their slightly lower contributions and their beauty. While they are not significantly more selfish, they are more beautiful, and returns to beauty are high. We return to the

---

[22]Men are found to have higher variance in behavior relative to women in other research. Andreoni and Vesterlund (2001) find that men are more likely to be completely selfish or completely altruistic. Pinker (2002) finds higher variance in the IQ scores of men relative to women.

interaction of beauty and gender in Section 5.3. First, we address the question of why men do better in the Information treatment than women.

5.2.2. "Good Examples" and Sex Stereotypes

Are people more cooperative with men than with women? Table 7 reports the results of linear regressions showing how the average individual contribution to the group is affected by the gender composition of the group.[23] The regressions include individual fixed effects.[24] Model 1 shows that, over both treatments, as the number of men in the group increases,[25] the average individual contribution of group members increases as well.[26] These effects are significant. This suggests that people cooperate more with men than with women, even though Table 6 showed that men give no more on average than women.

When we consider treatment effects, Model 2 shows that, with information, the cooperation effect of more men in the group increases two fold.[27] Without information, the effect is not significantly different from zero. That is, people are

---

[23] The dependent variable is the average amount a subject contributed to a group over the 8 rounds he played with that group. That is, the average that individual i contributes to group j is defined as $\overline{g}_{ij} = \sum_{t=1}^{8} g_{ijt}/8$. Each subject plays with 5 different groups, so in total, there are 400 observations for the 80 subjects.

[24] These are captured with identification dummy variables. Note that these variables also capture the treatment effects of information.

[25] This is the percent of men in the group other than the subject himself. So, for a group with three men and two women, each man would have 50% of the group as men, and each woman would have 75% of the group as men.

[26] The dependent variable, average contribution to the public good, is censored at zero and 20 since subjects are only endowed with 20 tokens in each round. Although the models in Table 7 were run as OLS linear regressions, very similar results hold if the models are run as a double-truncated Tobit.

[27] Across treatments, there is no significant difference in the distribution of the gender composition of groups. The p-value for a Fischer Exact test is 0.522. Thus, the significant treatment effects are *not* due to one treatment having more groups with more men than the other treatment.

more cooperative in groups with more men only when performance is observable.[28] For example, the average individual contribution to a group where half of the group, other than oneself, is comprised of men will be about 3 tokens higher in the Information treatment than in the No Information treatment.

Table 7
OLS Regression Coefficients
Dependent Variable: Average Contribution to the Group
Pooled Data

	Model 1	Model 2
Constant	6.36	8.41
	(0.0002)	(0.0001)
% Men in group other than self	3.07	-0.36
	(0.0011)	(0.7970)
% Men in group other than self * Info		6.06
		(0.0011)
Individual Fixed Effects	yes	yes
R^2	0.62	0.63

Note: p-values in parentheses, $n = 400$

Could this significant increase in cooperation with information be explained by favoritism based on sex stereotypes? If people expect attractive people to be more cooperative, could they also expect women to be more cooperative, especially since they are rated higher in attractiveness? From our ratings data on helpfulness, indeed, more people expect women to be helpful than men, but, perhaps more importantly, more people expect men to be more *unhelpful* than women.[29] Forty-one percent of men are rated as unhelpful, but only 10% of women are rated as such.

This would suggest that people are more cooperative as the number of men in the group increases because they discover that men are not as unhelpful as they

---

[28]Note that this effect also holds for both men and women separately.

[29]Twenty-five percent of women are rated in the upper quartile of helpfulness compared to 12% of men.

21

expected. So, even though men give no more on average than women, men may be able to engender more cooperation because they exceed expectations.

Perhaps a stronger explanation of men's ability to engender cooperation may come from the "good example" that they set. As noted earlier, men are more likely than women to contribute nothing or contribute their full endowment, but as the number of men in the group increases, the probability of contributing zero declines. A Tobit regression on the average probability of contributing zero in a group shows a significant decline in this probability as the percent of men in the group increases.[30] This holds for pooled data, for men only, and for women only.[31]

While the probability of contributing the "bad" extreme declines with the presence of more men, the probability of contributing the "good" extreme increases, but only for men. Table 8 shows how the average probability that a man will contribute the full endowment to the group changes as the percent of men in the group increases.[32] The results for women are similar, but insignificant. Model 1 shows that men are more likely to contribute all of their tokens as the percent of men in the group increases, and Model 2 shows that, when performance is known, men are almost twice as likely to contribute the full endowment.

---

[30]The dependent variable is the average probability that a subject contributed zero tokens to a group over the 8 rounds he played with that group. That is, the average probability that individual i contributes zero tokens to group j is defined as $\overline{p}_{ij} = \sum_{t=1}^{8} I(1|g_i = 0)_{ijt}/8$, where $I(\cdot)$ is an indicator function that equals 1 when the contribution was 0 tokens. The coefficient on the variable "% men in group other than self" is, $\widehat{\beta} = -0.24$ $(s.d. = 0.07)$.

[31]When performance is known, as in Model 2 in Table 8, the probability of free riding declines even further, but the effects are insignificant.

[32]The dependent variable is the average probability that a subject contributed his full endowment to a group over the 8 rounds he played with that group. That is, the average probability that individual i contributes his full endowment to group j is defined as $\overline{p}_{ij} = \sum_{t=1}^{8} I(1|g_i = 20)_{ijt}/8$, where $I(\cdot)$ is an indicator function that equals 1 when the contribution was 20 tokens.

Table 8
Tobit Regression Coefficients
Dependent Variable: Average Probability of Contributing Full Endowment
Men Only

	Model 1	Model 2
Constant	-0.29	0.04
	(0.1683)	(0.8543)
% Men in group other than self	0.38	-0.15
	(0.0221)	(0.5533)
% Men in group other than self * Info		0.87
		(0.0076)
Individual Fixed Effects	yes	yes
Log-likelihood	-87.20	-83.75

Note: p-values in parentheses, n=205

This suggests that the reason that men make more money when performance is known is because they put on a better performance, especially in the presence of other men. Moreover, their performance may engender more cooperation from others.[33] While men free ride more than women, they also set a "good example" more than women. As the composition of the group changes from women to men, both men and women free ride less, but men set the "good example" even more. This significantly increases the contributions of others in the group, and men make more money.

5.3. The Combined Effects of Beauty and Gender

Do the effects of beauty and gender hold when we control for them together? Table 9 shows the effects of beauty and gender on individual payoff in each group.[34] The

---

[33]This result is consistent with Camerer, Ho, Chong, and Weigelt's (2002) idea of "strategic teaching." A person, recognizing that others learn in a repeated game, would make choices to maximize long-run payoffs. This would suggest that men set a good example to ensure cooperation, and higher earnings, in the future.

[34]The dependent variable is the total payoff a subject earned in a group over the 8 rounds he played with that group. Each subject plays with 5 different groups, so in total, there are 400

regressions control for individual and group effects.[35]

<div align="center">

Table 9

OLS Regression Coefficients

Dependent Variable: Total Individual Payoff (sum across all 8 rounds) with a Group

</div>

	Pooled (n=400)	Women (n=195)	Men (n=205)
Constant	**3.98**	**4.33**	**4.13**
	(0.31)	(0.47)	(0.36)
Attractive	**0.58**	**0.62**	0.22
	(0.24)	(0.33)	(0.36)
Middle Attractive	**0.44**	0.22	**0.58**
	(0.18)	(0.27)	(0.24)
# Attractive people in group other than self	**0.53**	**0.55**	**0.49**
	(0.11)	(0.12)	(0.13)
% Men in group other than self	0.63	0.36	0.10
	(0.41)	(0.58)	(0.58)
Info	**0.91**	0.09	0.76
	(0.43)	(0.62)	(0.54)
Attractive*Info	-0.48	-0.42	-0.56
	(0.34)	(0.46)	(0.50)
Middle Attractive*Info	-0.38	-0.20	**-0.69**
	(0.24)	(0.35)	(0.38)
# Attractive people in group other than self*Info	**-0.66**	**-0.72**	**-0.50**
	(0.16)	(0.18)	(0.19)
% Men in group other than self * Info	0.62	**1.51**	**1.72**
	(0.57)	(0.79)	(0.82)
Individual & Group Random Effects	yes	yes	yes
$\overline{R}^2$	0.19	0.17	0.22

Note: standard errors in parentheses, and significant values (p-value < 10%) in bold.

---

observations for the 80 subjects.

[35]To address unobservable correlation across individuals and across individuals in the same group, the estimates use a two-way random effects error correction model.

In the pooled data, the beauty premium holds for attractive and middle attractive people, even when controlling for the beauty and gender composition of the group. Looking at men and women separately, it is attractive women that gain the beauty premium, but for men, only middle attractive men earn significantly more.

Interestingly, even when controlling for individual beauty, both men and women earn more money when the number of attractive people in the group increases. The premium to being in the presence of beauty, however, only holds when individual performance is unknown. Once the group knows what each individual member contributes, the premium disappears.

Table 10 shows the effects of beauty and gender on average contributions in each group.[36] As with the payoff regressions, these control for individual and group effects.

Looking at contributions in the pooled regression, we see that average contributions are higher in groups with more attractive people, but only when individual contributions are not revealed. Once individual contributions are revealed, people are less cooperative as the number of attractive people in the group increases. As we saw previously, this change in behavior is driven by men. We also see that people are more cooperative in groups with more men when contributions are known, confirming earlier results.

---

[36]The dependent variable is the average contribution a subject makes in a group over the 8 rounds he played with that group. Each subject plays with 5 different groups, so in total, there are 400 observations for the 80 subjects.

Table 10
OLS Regression Coefficients
Dependent Variable: Average Contribution to the Group

	Pooled (n=400)	Women (n=195)	Men (n=205)
Constant	**9.81**	**8.35**	**10.05**
	(1.60)	(2.29)	(1.89)
Attractive	1.11	0.29	4.01
	(1.63)	(2.25)	(2.47)
Middle Attractive	**-2.98**	-2.06	**-4.12**
	(1.29)	(1.92)	(1.75)
# Attractive people in group other than self	**1.05**	**1.07**	0.65
	(0.45)	(0.49)	(0.56)
% Men in group other than self	-2.51	-1.36	-0.73
	(1.94)	(2.37)	(2.35)
Info	-2.49	-2.12	-3.20
	(2.21)	(2.95)	(2.92)
Attractive*Info	-2.13	-3.35	-2.94
	(2.26)	(3.19)	(3.46)
Middle Attractive*Info	**3.48**	0.52	**5.75**
	(1.77)	(2.54)	(2.75)
# Attractive people in group other than self*Info	**-1.32**	-0.47	**-1.91**
	(0.64)	(0.70)	(0.76)
% Men in group other than self * Info	**7.20**	**7.21**	**7.17**
	(2.64)	(3.18)	(3.28)
Individual & Group Random Effects	yes	yes	yes
$\overline{R}^2$	0.17	0.18	0.26

Note: standard errors in parentheses, and significant values (p-value < 10%) in bold.

In sum, when controlling for individual beauty and gender and for the beauty and gender composition of the group, our main results still hold. There is a premium to beauty, but that premium disappears when performance is known. Others also benefit from being in the presence of beauty because people are more cooperative. This monetary return, however, disappear when group members can

clearly identify individual contributions. People are far less cooperative as the number of attractive people in the group increases. Being in a group with more men does increase average contributions, but only when individual contributions are observed.

6. Discussion and Conclusions

We used laboratory experiments to examine the returns to beauty and gender in a public goods game. Group members were identified with digital photographs. In one treatment, group members knew only the total amount contributed to the public good by the group, while in the other treatment they also knew the exact contribution of each group member. Subjects' photos were later independently rated in terms of physical attractiveness and how helpful-looking the subject appears.

In general, our results show significant and surprising effects of beauty and gender on earnings. When performance is unknown, people tend to reward beauty and females. And, when performance is known, the beauty premium disappears and the female premium switches to a male premium. These results appear to be rooted in beauty and sex stereotypes and have three main components.

First, relative beauty is rewarded. Attractive people make more money than middle attractive people, who in turn make more money than unattractive people. This premium to beauty mirrors Hamermesh and Biddle's (1994) results and can be quite large. Attractive people earn 15% more than unattractive people. This premium disappears, however, when actions are known. We find the beauty premium is not due to the actions of attractive people, but seems to be due to the expectations of how attractive people will behave, as suggested by psychologists'

27

social expectancy theory. People expect beautiful people to be more cooperative, and thus behave more cooperatively toward them when they are in the same group. But when information on decisions reveals that attractive people are no more or less cooperative than others, contributions decline relative to groups with no attractive people. This seems to reflect disappointed stereotypes of attractive people.

Second, women benefit for being stereotyped as helpful. Women make 7% more than men when contributions are not known. Looking more closely, we find the higher payoff enjoyed by women is not due to more selfishness on their part but, rather, may be a consequence of their beauty and sex stereotypes. Women are rated significantly higher in attractiveness than men, and their higher payoff seems to be the reward to their beauty. Also, because women are stereotyped to be more helpful than men, people are more cooperative in groups of women than in groups of men when performance is unknown.

Third, when performance is known, the gender premium switches from a female premium to a male premium. That is, while women gain the benefits of beauty, men gain the benefits of information. The payoff for men is 15% higher when performance is known. The payoff for women does not change. Indeed, this benefit to information for men is comparable in value to the attractiveness premium for women. Why do men attain this information benefit and women do not? With information on actions, groups with more men are able to engender more cooperation from others. That is, they set good examples that are followed by others. Free riding declines in groups with more men, and, in apparent male-competition, men are more likely to contribute their full endowment as the proportion of men in the group increases.

There are obvious implications of these results for experimental methodology. Indeed, this work justifies the extreme measures experimenters take to protect the identities of subjects. But are there implications of the beauty and gender effects we found for the real world?

The most striking result from this research is that, on average, men and women did not behave very differently, and attractive people did not behave appreciably differently than unattractive people. Nonetheless, beauty and gender had significant effects on earnings. This happened because beauty and gender affect the way people were treated by others.

Hamermesh and Biddle (1994) found that there is some sorting of beautiful people into occupations in which employers report that looks may increase productivity, such as in sales, but that this is not strong enough to explain much of the premium to beauty. This leaves social effects rather than productivity effects as the natural place to look for explanations. Our research suggests one effect of beauty is rooted in stereotypes. Beautiful people tend to be in more successful teams because other team members are more cooperative in the presence of beautiful people. This is true when effort is not observable, and suggests that a beauty premium may be more likely to exist when productivity is not perfectly observed, implying that beautiful people may also sort into occupations where individual productivity is difficult to measure. This hypothesis could be examined with labor market data.

The gender gap can also possibly be explained by a different stereotype for men. While beautiful people benefit from a stereotype of being more cooperative team members, men in our experiment benefit from exceeding the low expectations others have for them. People seem to expect men to be less helpful than they are.

When they see men exceeding expectations, they respond by following their lead. Men may be seen as better leaders despite the expectation that they are more selfish. In addition, when men are in groups together, they benefit from a kind of male-competitiveness for contributing to the public good. In the end, men are more likely to be in successful groups when information on contributions is known, and the more men in the group the greater the advantage to men. This again suggests hypotheses that can be tested with labor market data. Professions where effort can be verified more easily may favor men, and male-dominated professions may also have a bigger wage gap.

This paper illustrates the possibility for subtle stereotypes to have major economic impacts. However, economists are only beginning to consider seriously causes and consequences of stereotypes. Further work—including theory, policy and experiments—could fruitfully shed light on this important issue.

References

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

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

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

Bohnet, Iris and Bruno Frey (1999b), "Social Distance and Other-Regarding Behavior in Dictator Games: Comment," *American Economic Review*, 89(1), 335-339.

Bosman, Ronald and Frans van Winden (2002), "Emotional Hazard in a Power-to-Take Experiment," *The Economic Journal*, 112, 147-169.

Buchan, Nancy, Rachel Croson and Sara Sonick (2003), "Trust and Gender: An Examination of Behavior, Biases, and Beliefs in the Investment Game," Working Paper, The Wharton School, University of Pennsylvania.

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

Camerer, Colin, Teck Ho, Juin-Kuan Chong, and Keith Weigelt (2002), "Strategic Teaching and Equilibrium Models of Repeated Trust and Entry Games," Working Paper, California Institute of Technology.

Cortina, J.M. (1993), "What is Coefficient Alpha? An Examination of Theory and Applications," *Journal of Applied Psychology*, 78, 98-104.

Darley, J.M. and R.H. Fazio (1980), "Expectancy Confirmation Processes Arising in the Social Interaction Sequence," *American Psychologist*, 35, 867-881.

Devine, Patricia (1989), "Stereotypes and prejudice: Their automatic and controlled components," *Journal of Personality and Social Psychology*, 56(1), 5-18.

Eckel, Catherine and Philip Grossman (1998), "Are Women Less Selfish Than Men? Evidence from Dictator Experiments," *Economic Journal*, 108(448), 726-35.

Eckel, Catherine and Phillip Grossman (2000), "Differences in the Economic Decisions of Men and Women: Experimental Evidence," *Handbook of Experimental Economic Results,* forthcoming.

Eckel, Catherine and Rick Wilson (2003), "Is There a Mechanism for Detecting Trustworthiness?" Working Paper, Virginia Tech.

Fershtman, Chaim and Uri Gneezy, "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, 116(1), 351-377.

Glaeser, Edward, David Laibson, Jose Scheinkman, and Christine Soutter (2000), "Measuring Trust," *The Quarterly Journal of Economics*, 115(3), 811-846.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), "Performance in Competitive Environments: Gender Differences," *Quarterly Journal of Economics*, 118(3), 1049-1074.

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

Langlois, Judith, Lisa Klakanis, Adam Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot (2000), "Maxims or Myths of Beauty? A Meta-Analysis and Theoretical Review," *Psychological Bulletin*, 126(3), 390-423.

McPherson, Miller, Lynne Smith-Lovin, James M. Cook (2001), "Birds of a Feather:

Homophily in Social Networks," *Annual Review of Sociology*, 27, 415-444.

Mobius, Markus and Tanya Rosenblat (2003), "Why Beauty Matters," Working Paper, Wesleyan University, forthcoming *American Economic Review.*

Mulford, Matthew, John Orbell, Catherine Shatto, and Jean Stockard (1998), "Physical Attractiveness, Opportunity, and Success in Everyday Exchange," *American Journal of Sociology*, 103(6), 1565-1592.

O'Neill, June (2003), "The Gender Gap in Wages, circa 2000," *American Economic Review*, 93(2), 309-314.

O'Neill, June (1998), "The Trend in the Male-Female Wage Gap in the United States," Ferber, Marianne, ed. *Women in the Labour Market. Volume 1*, Elgar Reference Collection. International Library of Critical Writings in Economics, 90, 520-45.

Pinker, Steven (2002), *The Blank Slate: The Modern Denial of Human Nature*, New York: Viking Press.

SAS Institute Inc. (1988), *SAS Procedures Guide: Release 6.03 Edition*, Cary, NC: SAS Institute, Inc.

Scharleman, Jorn, Catherine Eckel, Alex Kacelnik, and Rick Wilson (2001), "The Value of a Smile: Game Theory with a Human Face," *Journal of Economic Psychology*, 22(5), 617-640 .

Solnick, Sara and Maurice Schweitzer (1999), "The Influence of Physical Appearance and Gender on Ultimatum Game Decisions," *Organizational Behavior and Human Decision Processes*, 79(3), 199-215.

Zebrowitz, L.A. (1997), *Reading Faces: Window to the Soul?*, Boulder, CO: Westview Press.