When Are Appearances Deceiving? The Nature of the Beauty Premium

Deryugina, Tatyana

August 2013

When are appearances deceiving? The nature of the beauty premium

First Draft: December 2012

Current Draft: August 2013

Abstract:

We design a laboratory experiment to illuminate the channels through which relatively more attractive individuals receive higher wages. Specifically, we are able to distinguish taste-based discrimination from rational statistical discrimination and biased beliefs. Using three realistic worker tasks to increase the external validity of our results, we find that the "beauty premium" is highly task-specific: while relatively more attractive workers receive higher wage bids in a bargaining task, there is no such premium in either an analytical task or a data entry task. The premium in the bargaining task is driven by biased beliefs about worker performance. We find that there is substantial learning after worker-specific performance information is revealed, highlighting the importance of accounting for longer-run interactions in studies of discrimination.

JEL Classification: C91, J31, J71
Key Words: beauty premium, discrimination, economic experiments, labor markets

1. Introduction

Labor market discrimination based on characteristics such as gender, age, race, and national origin is illegal. Appearance-based discrimination, while not currently unlawful in most labor markets, has been the subject of several lawsuits in recent years.[1] In the academic literature, several studies have found that people who are relatively more attractive are paid more, even when the situation does not appear to warrant it (Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998). This phenomenon has been termed the "beauty premium." It appears to be pervasive: versions of the beauty premium have been found in settings that include credit markets (Ravina, 2009), professional sports (Berri et al., 2011), and even elections (Berggren et al., 2010). Moreover, there is extensive evidence that beauty is correlated with career choices, including the choice to become a criminal (Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998; Mocan and Tekin, 2010; von Bose, 2012; Deryugina and Shurchkov, 2013).

One potential explanation for the beauty premium is that appearance may in fact be positively correlated with skills that are important for job performance but are not easily observed and thus cannot be controlled for in an empirical study, such as the ability to be persuasive. Another is that employers may overestimate the skills of relatively attractive people simply because they're attractive. Finally, employers may have unbiased beliefs about performance but prefer hiring more attractive people ("taste-based discrimination"). Because most studies are observational rather than experimental, the existing evidence does not allow for these three channels to be credibly separated.

We perform a computer-based laboratory experiment that allows us to distinguish these three causes from one another.[2] We estimate the extent to which employers *correctly* predict the relationship between appearance and task performance, thus determining what share of the beauty premium, as measured by the wage bids, is statistical discrimination. Then, by appropriately controlling for performance predictions, we are able to estimate the portion of the beauty premium that is not driven by performance expectations. The direct estimation of this taste-based component of the beauty premium also adds to previous work in which researchers were able to impute some types of taste-based discrimination only from the way in which employer beliefs are elicited. Finally, because we observe workers' actual performance, we also estimate the correlation between performance and worker appearance. Together with the relationship between employer performance predictions and worker appearance, this allows us to identify any biased beliefs about the skills of comparatively attractive people.

Another innovation in our study meant to capture the diversity of real-world labor markets is that we estimate the extent to which the beauty premium is context-specific. We do this by randomly assigning one of three tasks to workers: a data entry task, an analytical task, or a bargaining task in which workers see pictures of their bargaining opponents. The rest of the experimental procedure remains identical across the tasks. To our knowledge, our study is the first to credibly test whether the beauty premium varies with the types of skills involved in completing a task and, if so, to determine why. Because the tasks we choose are also more realistic than previously studied tasks, the variation we observe is likely to be similar to the variation in the beauty premium outside of the laboratory, therefore increasing the external validity of our results.

[1] See for example Yanowitz v. L'Oreal USA, Inc. (2005) and Brice v. Resch and Krueger Int'l, Inc. (Corbett, 2011).
[2] See Mobius and Rosenblat (2006) for more experimental evidence on the beauty premium in the labor market, Andreoni and Petrie (2008) in public goods games, and Wilson and Eckel (2006) in trust games.

Our study is also the first to investigate how the signal value of appearance regarding a worker's ability changes with information. It is possible that attractiveness is used as a proxy for ability at the "recruiting" stage, modeled in our experimental setting with the first round. However, it might become increasingly irrelevant as employers observe actual worker performance. To test for the existence of this type of learning, we reveal workers' first round performance to all employers. We then repeat the prediction, bidding, and task performance stages, allowing employers to update their bids and expectations. We then estimate what portion of the beauty premium disappears once performance measures for each worker are available.

We find a significant beauty premium in bargaining but not in the other two tasks (data analysis or data entry). In particular, a one-standard-deviation increase in worker attractiveness is associated with a 28 per cent increase in the employer's wage offer when the workers engage in a bargaining task, even after including extensive controls. By dividing attractiveness ratings into quintiles, we show that the most attractive subjects (those in the top quintile of beauty) benefit most significantly relative to the least attractive subjects. Our conclusion that the beauty premium is highly context-specific is consistent with results reported in the non-experimental literature, which finds substantial beauty-based sorting into different occupations (Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998; Mocan and Tekin, 2010; von Bose, 2012; Deryugina and Shurchkov, 2013).

The beauty premium is completely explained by statistical discrimination: employers believe that more attractive workers will perform better in bargaining, where workers can see one another's picture, but not in data entry or data analysis. This belief turns out to be incorrect: there is no significant relationship between the attractiveness rating and performance in any of the tasks, including bargaining.

We find that the beauty premium completely vanishes in the second round of bidding, which suggests that employers learn quickly that performance is uncorrelated with attractiveness. Past performance is also a significant determinant of wages in the second round, largely because it affects employer beliefs about worker performance. Both these facts suggest that there is substantial updating by employers and that biased beliefs correct themselves quickly when objective information about performance is available.

The intersection of gender and the beauty premium has been a topic of several non-labor market studies (Solnick and Schweitzer, 1999; Rosenblat, 2008). We build upon this literature to find no systematic evidence that female or male workers benefit disproportionally from attractiveness. Comparatively attractive males are expected to perform significantly better in bargaining, but do not receive larger wage bids than comparatively attractive females or less attractive males. Moreover, this expectation turns out to be incorrect.

Our general approach shows that having measures of both (a) expectations about a worker's performance and (b) the willingness to hire the worker (by giving her a relatively higher wage offer) are helpful for separating different kinds of discrimination. This idea can be applied in other settings and to other types of discrimination. For example, asking car dealers how much money they expect to make on a particular sale *and* observing their actual bargaining behavior can help determine whether offers made to women and minorities are driven by statistical or taste-based discrimination, something an earlier study on the subject was not able to determine conclusively (Ayres and Siegelman, 1995). Although other methods for distinguishing taste-based from statistical discrimination have been used, direct elicitation of beliefs may be preferable, as it relies on fewer assumptions.[3]

Because of the experimental nature of our work, addressing its external validity is important. According to the theory of discrimination (Becker, 1957), taste-based discrimination is predicted to arise in real-world settings when employers expect to derive direct utility from future face-to-face interactions with relatively more attractive workers. Such face-to-face interactions are absent in our experiment, making it difficult to claim that our findings about the absence of taste-based discrimination generalize to the typical office setting. However, the importance of settings where appearance is observable only through a photograph is growing with the rise of online labor markets, such as oDesk (Pallais, 2013), and online credit markets, such as Prosper (Ravina, 2009; Duarte et al., 2012). Moreover, with the increasing number of "telecommuting" workers, the old paradigm of "face time" is changing.[4] In addition, hiring in some cases is done by temp agencies or human resources departments, which are functionally removed from daily interactions with the hired workers. Our findings about taste-based discrimination are most directly relevant to such settings.

The remainder of the paper is organized as follows. In Section 2, we present an overview of our experimental procedures and descriptive statistics. Section 3 outlines the framework that allows us to differentiate biased beliefs about performance, statistical discrimination, and taste-based discrimination. Section 4 reports and discusses the results, and Section 5 concludes.

2. Overview of the Experiment

2.1. The Stylized Labor Market

The experiment was conducted at the Decision Science Laboratory at Harvard University. Subjects were undergraduate and graduate students from Harvard and other Boston-area universities. Each session included four employers and four workers. Sessions differed according to the task workers had to perform, which was randomly assigned for each session (see Section IIB below). The first four subjects to arrive at the laboratory and sign the consent form (our "employers") were immediately taken from the waiting room, photographed, and seated at their stations. The next four subjects to arrive at the laboratory and sign a consent form (our "workers") were photographed and seated afterwards.[5] In order to avoid further face-to-face interactions between the two groups, employers and workers were seated at stations separated by a wall divider.

All subjects started by having their photograph taken and answering survey questions about several characteristics that are relevant to the labor market (student status, major, and GPA as well as levels of typing, analytical, and communication skills) before being told whether they would be employers or workers. After receiving the experimental instructions, which included detailed information about the task workers would perform, employers were granted access to a

[3] Evidence consistent with statistical and taste-based discrimination has been found by Castillo et al. (2012) in a field study on gender differences in bargaining outcomes over taxi fares in Peru. Their study differs from ours insofar as it takes place in a non-labor-market setting in which beliefs are not elicited directly, but rather are inferred from observed initial price quotes.

[4] For example, the online marketplace, oDesk, consists of workers all over the world who complete approximately 200,000 hours of work per week remotely (Pallais, 2013). In the US, telecommuting increased 73% from 2005 to 2011, and 64 million U.S employees holds a job that is compatible at least part-time telework (Global Workplace Analytics, 2011).

[5] Most subjects arrived in the laboratory within ten minutes of one another, which ensures almost random role assignment. By assigning the role of employer to the first four subjects and removing them from the waiting room, we minimized the likelihood of face-to-face interactions between employers and workers that may have otherwise occurred during the initial waiting period.

website that displayed worker photographs and the corresponding "résumés" based on each worker's survey answers. In 25 sessions, photos were shown on the front webpage with links to résumé information underneath each photo. In 22 sessions, this order was reversed.

The remainder of the experiment, programmed using the standard zTree software package (Fischbacher, 2007), consisted of two procedurally identical rounds. Each round started with a prediction stage during which employers submitted estimates for the expected performance of each worker in the subsequent task (E_{ij}), where i indexes employers and j indexes workers, and workers submitted estimates for their own expected performance (E_j). This information was kept secret from all other subjects. The wages of both employers and workers were partly determined by the accuracy of their predictions, ensuring that subjects had incentives to guess correctly.

Next, employers submitted wage offers to "hire" workers. The total amount offered to four workers could not exceed a predetermined maximum number of points.[6] We employed a second-price sealed-bid auction to allocate workers to employers: the employer with the highest wage offer for a particular worker "hired" that person and had to pay the worker the second highest wage (W_j) offered to that worker. Each employer could be matched with between zero and four workers, depending on the wage offers. A worker could be left unmatched if all four employers offered a zero wage to that worker, although this did not happen in practice. The wage amount (if any) was not revealed to the worker until after the task completion stage.[7] The identity of the employer was never revealed to the worker. Employers had full knowledge about the tasks workers were to perform prior to making performance predictions and wage bids.

The task completion stage began after employer–worker matching was established. The task was randomly chosen prior to the start of the session to be a bargaining task, a data entry task, or a data analysis task (see detailed task descriptions below). Table 1 shows the number of sessions for each task type and the corresponding number of subjects who participated in a given session.

[TABLE 1 ABOUT HERE]

Each round ended with an information screen. Employers learned about the performance of every worker and their own payoffs for the round. Workers learned about their own performance and payoffs for the round, including any wage payment. The following equations represent the total within-round payoffs.

Employer i's Payoff:

$$\pi_i = 125 + \frac{1}{3}\sum_{j=1}^{4} P_t Y_j \times Hire_{i,j} - \sum_{j=1}^{4} W_j \times Hire_{i,j} - M_t \sum_{j=1}^{4} |Y_j - E_{i,j}|$$

Worker j's Payoff:

[6] In 22 earlier sessions, this amount equaled the employer's endowment of 125 points, while in the 25 subsequent sessions this amount was raised to 175 points with the endowment remaining at 125 points. The increase was meant to allow employers to base their bids on their estimates of expected worker performance rather than on the mechanical constraint imposed by the bid maximum. The bid maximum does not affect the results; see the online appendix for details.

[7] If employers expect more attractive workers to be more likely to reciprocate a higher wage with higher effort, then, by withholding the wage offer information until after the worker completes the task, we are shutting down the potential "gift-exchange" mechanism behind the beauty premium. First, introducing this additional channel would greatly complicate our already complex design. Furthermore, the gift-exchange channel is unlikely to drive the beauty premium, as previous work has found that more attractive individuals do not exhibit greater levels of reciprocity relative to their less attractive counterparts (see Wilson and Eckel (2006) in trust games, for example).

$$\pi_j = 25 + \frac{2}{3}P_t Y_j + W_j - M_t|Y_j - E_j|$$

Where $i \in \{1,4\}$ is the set of employers, $j \in \{1,4\}$ is the set of workers, and $t \in \{Data\ Entry, Data\ Analysis, Bargaining\}$ is the set of tasks; P_t is the piece rate of 5 points for t = *Data Analysis* and 1 point for the other tasks; M_t is the weight on the deviation of the performance estimate from actual output and equals $\frac{5}{4}$ for t = *Data Analysis* and $\frac{1}{4}$ otherwise; $Hire_{i,j}$ is an indicator function that takes on the value of 1 if worker j was hired by employer i, and 0 otherwise. The last term in both equations represents a "misprediction penalty" that we include in order to incentivize truth-telling in accordance with other studies (Mobius and Rosenblat, 2006).[8]

At the end of the session, all subjects filled out a post-experiment questionnaire that asked for detailed demographic information. Mean earnings in the experiment (including the show-up fee) equaled \$17.12 with a standard deviation of \$2.22. Sessions lasted approximately one hour. Experiment instructions and questionnaire contents are available in the online appendix.

2.2. *The Tasks*

We deliberately focus on tasks with which employers are more likely to be familiar and thus in which appearance-based differences in expectations are more likely to be correct. The abovementioned laboratory study on the beauty premium (Mobius and Rosenblat, 2006) focuses on a task type with which employers are unlikely to have prior experience and finds a substantial beauty premium despite the absence of beauty-based differences in performance. Because of the lack of familiarity with the task, however, it is not clear whether it is reasonable for employers to expect appearance-based performance differences. For example, it is possible that employers are extrapolating their beliefs from other situations in which more attractive people do have a performance advantage. Having workers perform realistic tasks increases our confidence in determining whether any beauty-based differences in performance predictions or wage bids are due to rational performance expectations, biased beliefs, or tastes.

Ex-ante, employers may have differing expectations about the relationship between attractiveness and performance in each of the three tasks, either because of a true correlation between the two or because of biased beliefs. For example, elementary school children presented with photos of 10 individuals (all scientists) "showed a decided tendency to identify the smiling pictures as not being scientists" (Bottomley et al., 2001). In a "Draw a Scientist" experiment, children typically draw an unattractive white male wearing a white lab coat and glasses (Chambers, 1983). In a recent study, Deryugina and Shurchkov (2013) find that comparatively attractive female undergraduates perform worse than their less attractive counterparts on blindly graded quantitative reasoning tests and SATs and are less likely to choose a science major or become scientists. Thus, in the more difficult analytical task, we may find comparatively attractive individuals receiving lower wage offers because of the common stereotype that people who are good at such tasks are less attractive.

[8] Incentivized belief elicitation may distort worker incentives during the task completion stage, leading to a "hedging bias". On the other hand, monetary incentives increase truth-telling and reduce the "noise" in the beliefs data (Gachter and Renner, 2006). We prioritize the latter issue, given the recent finding that the former may not be a serious concern in belief elicitation experiments (Blanco et al, 2010). In addition, to minimize concerns about hedging bias, we chose a relatively small M_t and a generous exchange rate from points to money to ensure a salient reward for any additional effort exerted once the expected predicted performance level has been attained.

On the other hand, researchers have found that attractive subjects get higher offers and therefore outperform their less attractive counterparts in simple bargaining games even without face-to-face interactions, such as the ultimatum game (Solnick and Schweitzer, 1999). Thus, employers in our experiment may expect more attractive workers to perform better in our bargaining task, and therefore incorporate these expectations in their wage offers. Finally, in the simple data entry task, we do not expect to see beauty-based performance or wage offer differences unless attractiveness happens to be correlated with another previously unidentified skill.

2.2.1. Data Entry

In the data entry task, workers had six minutes to enter numerical data that they read off a sheet of paper into an Excel spreadsheet. The goal was to enter as much data as possible. The data consisted of various economic statistics for regions in Russia. The spreadsheets had been opened on the workers' computers prior to the start of the experiment with the column and row headings prepared in advance, so that subjects had only to enter numerical values into the correct cells. The data had to be entered exactly as it appeared to receive credit. Workers were credited with one point per correctly entered item. There was no penalty for an incorrectly entered item.

2.2.2. Data Analysis

In the data analysis task, workers answered as many mathematical questions as possible, up to a maximum of 30 questions. Questions were based on data that were similar to those used in the data entry task. Workers had six minutes for the first 15 questions and six minutes for the second 15 questions. Because some questions required basic mathematical calculations, workers could use calculators that had been placed on their desks in advance. Workers were credited with five points per correctly answered question, and there was no penalty for answering questions incorrectly. In our analysis, we likewise convert performance measures into points, multiplying the number of answers by five points.

2.2.3. Bargaining

In the bargaining task, workers were randomly assigned as buyers or sellers of a "widget" and participated in three 90-second periods of a standard double-auction. Including the time it took workers to read the information screen, which was not part of the 90-second limit, the bargaining task lasted about six minutes, on average. Workers were randomly re-matched and roles were randomly assigned with every new bargaining period. Each worker saw a photo of his or her bargaining partner on a computer screen.[9] Every time a transaction was made, the seller's profit equaled the difference between the price and the seller's true cost of the "widget," and the buyer's profit equaled the difference between the buyer's true value and the price of the "widget." Profits were calculated in tokens and then converted into points at the rate of 1 token = 1 point. If the time ran out before a transaction was made, both the buyer and the seller earned 0 tokens in that bargaining period. Each token was equivalent to one point for the purposes of calculating the total payoff for the round. Buyers' values and sellers' costs were determined randomly from two uniform distributions. In some cases, the buyer's value was below the

[9] Importantly, employers were made fully aware that workers could see the photo of their bargaining partner, but that no face-to-face interactions among workers would take place at any point.

seller's cost, making profitable agreements impossible. To avoid the possibility of negative profits, sellers could not agree to an offer that was lower than their cost and buyers could not agree to an offer that was higher than their value.

2.3. The Rating Procedures

The rating portion of the experiment was conducted at the University of Illinois, Urbana-Champaign (UIUC). Subjects were undergraduate and graduate students from UIUC. During each session, 4–15 subjects (raters) were instructed to view and evaluate photos on a scale from 1 (homely) to 10 (strikingly handsome or beautiful).[10] Each rater was asked to look through four sets of 100 photos, which appeared in random order within each photo set. Due to the large number of photos, each rater evaluated only a subset of photos. The individual rating variable used in subsequent analysis is demeaned by the rater's average across the photos that appeared in the same photo set; in other words, rater by photo-set fixed effects are implicitly controlled for in our analysis.

Each rating session lasted between forty minutes and one hour, including the reading of the instructions and payment. Raters were paid a show-up fee of \$5 and an additional \$7 payment for completing the task of rating all photos and providing demographic information.[11]

2.4. Descriptive Statistics

We start with a total of 376 subjects split evenly between employers and workers. Our main unit of observation is employer-worker pairs, of which we have 16 per session (four employers each bidding on four workers) for a total of 752 pairs. However, a few subjects drop out of our sample. First, we exclude two pilot sessions held on December 7, 2011 from the main analysis. Second, we drop a subject who self-identified as a non-student (an employer). Third, we drop a subject who participated in our experiment twice, keeping the first instance (an employer) and dropping the second instance (a worker). Fourth, we drop five employers who did not use the worker résumé information (that is, who did not click on the worker's photo or résumé). Including these observations does not significantly change the results. Finally, we also drop two subjects (workers) who chose to withdraw from our study after the experiment. The final dataset consists of 174 employers and 177 workers for a total of 685 employer-worker pairs.

Table 2 provides the summary statistics for employers and workers by task, gender, and round. The employers' outcomes across our three tasks appear comparable. In Round 1, the average wage bid ranges from 24 to 29. While the mean wage bid in data analysis is statistically lower than in data entry or bargaining (with t-test p-values of 0.043 and 0.026, respectively), the difference of around four points on average is relatively small. Similarly, the employers predict statistically lower performance in data analysis (in points) relative to the other two tasks (t-test p-value of < 0.001), but the differences are modest in size.

Female employers bid slightly higher than males, but the differences are not statistically significant. Similarly, there are no significant gender differences in employers' performance predictions and earnings in any of the tasks.

[10] The scale was expanded from a 1–5 point scale previously used in the literature (Hamermesh and Biddle 1994, 1998) to a 1–10 point scale.
[11] Summary statistics for rater demographic information can be found in the online appendix.

[TABLE 2 ABOUT HERE]

In Round 1, male workers predict that they will perform better than female workers in the bargaining task (t-test p-value of 0.076). These predictions turn out to be correct, with males outperforming females by about 24 points in this task.

Workers have significantly lower payoffs in data analysis than in the other two tasks in the first round (t-test p-value < 0.001). Workers earn significantly higher payoffs than employers in data entry and bargaining (t-test p-values of < 0.001 and 0.026, respectively), but employers earn significantly greater payoffs in data analysis (t-test p-value of 0.009).

In Round 2, the average wage bid ranges from 22 to 33. Females placed higher bids than males in the bargaining task (t-test p-value of 0.077). Worker performance predictions in the second round do not significantly differ by gender. There is no significant difference between male and female worker performance in data entry or bargaining, but there is a significant difference between males and females in data analysis, with males outperforming females by about 2.6 questions.

Again, payoffs for workers are significantly lower in data analysis than in the other two tasks (t-test p-value of < 0.001). Workers also make significantly higher payoffs than employers in data entry and bargaining (t-test p-values of < 0.001 and 0.030, respectively). Female employers earn significantly greater payoffs than female workers in data analysis (t-test p-value of < 0.001), but there is no difference between male employers and workers.

Table 3 provides the summary statistics on subject attractiveness and other characteristics by gender for employers and workers. There are no statistically significant differences in attractiveness between either men and women or employers and workers in our sample. In our sample of workers, there are no statistically significant differences in any of the individual characteristics, on average.

[TABLE 3 ABOUT HERE]

Before proceeding with more formal regression analysis, we estimate a simple correlation to test for the existence of the beauty premium in our experiment. Table 4 reveals the correlation between wage bids and attractiveness, without controlling for worker characteristics. When pooling all tasks, we find a significant positive correlation between the natural logarithm of the wage bid and worker attractiveness in both rounds. All specifications in Table 4 include date fixed effects with standard errors clustered by employer.

[TABLE 4 ABOUT HERE]

The results shown in columns 1 and 4 pool the data across tasks and include task fixed effects. Panel 1 shows that, on average, a one standard deviation increase in attractiveness increases the wage bid by about 16 per cent in both rounds. When we further decompose our analysis by task, we do not find any statistically significant relationship between beauty and wage bids in either the data entry or the data analysis task (Columns 2, 3, 6, and 7). However, the beauty premium is statistically significant in the bargaining task (Columns 4 and 8): a one standard deviation increase in attractiveness increases the wage bid by about 23 per cent in each round. Later in the paper, we elucidate the mechanisms behind this task-specific correlation.

In Panel 2 of Table 4, we investigate whether gender differences are more pronounced when we allow the beauty premium to vary for male and female workers. Columns 1–4 show that the beauty premium in the first round is driven by the higher wages offered to comparatively attractive female workers. However, we cannot reject the hypothesis that the coefficients on the male and female ratings are identical. In the second round, attractiveness is marginally significant for men in bargaining, although again we fail to reject the hypothesis that the coefficients on the male and female ratings are equal. In data entry, more attractive women receive significantly higher wage bids than men (the coefficients on the male and female ratings are statistically different), which is possibly due to females reporting higher typing skills than males.

3. Empirical Strategy

In this section, we outline a general framework for separating statistical discrimination from taste-based discrimination and testing whether statistical discrimination, if there is any, is based on rational or biased beliefs.

To separate taste-based discrimination from statistical discrimination, we use the fact that the performance prediction captures the employer's beliefs about actual worker performance, while the bid captures the value the employer derives from worker performance *and* from his or her attractiveness. Thus, if only statistical discrimination is present (whether or not beliefs about performance are correct), then any effect of attractiveness on wage bids should operate only through the performance expectation. In other words, once we properly control for the performance prediction, worker's attractiveness should have no further explanatory power in the case of pure statistical discrimination. If the effect of attractiveness on the employer's wage bid is significant after controlling for the performance prediction, we conclude that there is taste-based discrimination, as a result of which employers bid more on more attractive workers even though they do not expect them to be more productive. Because the employer's bid for such a worker is likely to exhibit a nonlinear relationship with her performance prediction, we allow the performance prediction to enter the specification flexibly, as the within-employer rank of the worker's performance prediction.[12]

$$\log(1+W_{ij}) = \beta a_j + \sum_{r=2}^{4} \gamma_r 1[Rank_{ij}(E_i[\theta_j]) = r] + X_j'\rho + \delta_T + \sigma_t + \varepsilon_{ij} \qquad (1)$$

where i indexes the employer and j indexes the worker. We suppress the round subscripts for tractability reasons. The variable W_{ij} is the bid of employer i on worker j in round 1 or round 2. The attractiveness rating is given by a_j, and $E_i[\theta_j]$ is employer i's expectation of worker j's performance. Worker characteristics are captured by X_j' and include indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. The coefficients of interest are represented by β, which captures taste-based discrimination, and the set γ_r, which captures statistical discrimination. The variable r indexes the performance prediction rank of the worker. In all specifications, we include a set of task fixed effects whenever we combine multiple tasks in a single regression (δ_T), as well as a set of date fixed effects (σ_t). Standard errors in equation (1) are clustered by employer.

[12] The optimal bidding strategy in our setting is not analytically tractable and is likely to vary nonlinearly with the performance prediction. Moreover, prior experimental literature finds that behavior consistently deviates from rational bidding strategies (e.g., Cooper and Fang, 2008).

In order to determine whether employer beliefs about worker performance in a given task are correct on average, we test whether actual performance θ_j and attractiveness α_j are correlated using the following specification:

$$\log(1 + P_{jT}) = \rho a_j + X_j' \delta + \delta_T + \sigma_t + u_{ij} \qquad (2)$$

where j indexes the worker and T indexes the task. We suppress the round subscripts for tractability reasons. The variable P_{jT} is the performance of worker j in task T in round 1 or in round 2. Worker characteristics are captured by X_j'. The coefficients of interest are represented by ρ, which captures the correlation between ability and attractiveness. In all specifications, we include a set of task fixed effects whenever we combine multiple tasks in a single regression (δ_T), as well as a set of date fixed effects (σ_t). Standard errors in equation (2) are clustered by worker.

Our design also allows us to examine the effect of information on the beauty premium. The two-round setting captures the way in which repeated interactions between employers and workers in the labor market increase the amount of available information over time. The first round can be thought of as a trial period, during which the employer has limited information to use in forming a belief about the worker's future productivity. A more precise signal indicating the worker's ability arrives later on, once past performance can be observed (for example, when the worker comes up for a review or renegotiates her contract). If attractiveness a_j is used as a signal of ability θ_j, it should be less informative in the second round, after a worker's actual performance is revealed. To test for this, we estimate equation (1) separately in round 1 and round 2. Moreover, we include the worker's past performance in the round 2 regressions as an explicit test of whether employers update their beliefs and bidding behavior based on new performance information.

Finally, we extend the analysis in equations (1) and (2) as follows. First, we allow the coefficient on the attractiveness rating to vary by attractiveness quintile to test for non-linear effects. Second, we separate the effects by gender to test whether the beauty premium may vary across male and female workers.

4. Results

4.1. The Sources of the Beauty Premium in the First Round

We begin our analysis with the first round. The first round represents an environment with limited information in which prospective employers make predictions about worker performance based on worker photos and résumé characteristics. Because we can observe these characteristics perfectly, we can test whether the information from the résumé helps to explain the correlation between attractiveness and wage bids we find in the absence of these controls.

Result 1: There is a significant beauty premium in the bargaining task in the first round. Furthermore, the beauty premium in bargaining is largest for the most attractive workers (those in the top attractiveness quintile). In the other two tasks, the beauty premium is absent on average, but exists for the moderately attractive workers in data entry.

Support for Result 1 comes from Table 5, which shows the relationship between the natural logarithm of the wage bid in round 1 and worker attractiveness, conditional on worker résumé characteristics: indicators for student status (graduate or undergraduate), major, self-reported abilities (typing, analytical, and communications), race, and gender. All specifications in Table 5 also include date fixed effects with standard errors clustered by employer.[13] Columns 1 and 2 pool the data across tasks and include task fixed effects. Column 1 estimates the simple relationship between the wage bid and the attractiveness rating, while Column 2 breaks the attractiveness rating into quintiles to allow for a potential nonlinear relationship between beauty and wage offers. On average, a one standard deviation increase in attractiveness leads to a 14 per cent increase in the wage offer. Relative to the bottom quintile, workers in all four of the top quintiles receive higher wage bids, but the significant positive effect appears only for workers whose beauty rating falls into the 4th quintile (above-average looks, but not the most attractive).[14]

[TABLE 5 ABOUT HERE]

When we further decompose our analysis, we observe that the beauty premium varies by task. In particular, on average, we do not find a significant effect of beauty on wage bids in either the data entry or the data analysis task (Columns 3 and 5). In data entry, workers in the 2nd, 3rd, and 4th attractiveness quintile receive significantly higher wage bids relative to the bottom quintile (Column 4), with the moderately attractive workers (4th quintile) receiving the largest premium.[15] On the other hand, none of the beauty quintiles receives a significantly higher wage than the first quintile in data analysis (Column 6). However, the coefficients are jointly significantly different from one another. For instance, moderately attractive workers (4th quintile) receive significantly higher wages than workers with below-average looks (2nd quintile), as well as those in the top attractiveness quintile (f-test p-values of 0.014 and 0.053, respectively).

On average, we find a beauty premium only in the bargaining task (Column 7): a one standard deviation increase in attractiveness increases the wage bid by 28 per cent in the bargaining task. Column 8 shows that the beauty premium in the bargaining task is strongest for the top quintile (the most-attractive workers).[16] We obtain similar results if we rank workers by attractiveness out of four in each session instead of using a continuous attractiveness variable: the beauty premium is still present only in bargaining with a unit increase in rank associated with an increase in the wage bid of 14% (p-value of 0.058). The rank results can be found in the online appendix.

The fact that the beauty premium shows up most consistently in the task that was expected to be "beauty related" ex-ante but not in the tasks that were expected to be "beauty unrelated" suggests that it is performance expectations, rather than tastes, that explain the existence of the overall beauty premium. Because employer expectations about the relationship between attractiveness and performance should play a role in the relationship between attractiveness and

[13] The results are robust to including employer fixed effects. The results are also not substantively different when we control for whether the workers' photos or résumés were shown to the employers first. We do not include these specifications in the paper, but estimates are available upon request.

[14] An f-test reveals that the coefficients on the 4th and 5th quintiles are not statistically significant from one another. Jointly, the two coefficients are statistically significantly different from zero (f-test p-value of 0.09). Furthermore, the coefficients on the 4th and 5th quintiles are jointly significantly different from the coefficient on the 2nd quintile (f-test p-value of 0.097).

[15] An f-test shows that the three coefficients are marginally jointly significant (p-value of 0.105) and that the difference between the 4th and the 5th attractiveness quintiles is statistically significant (p-value of 0.029).

[16] An f-test reveals that the coefficient on the 5th quintile is also statistically significantly different from the coefficient on the 2nd quintile (f-test p-value of 0.070).

wage offers, we next examine whether employers believe that more attractive workers are more productive in the three tasks.

Result 2: Employers expect more attractive workers to be more productive in the bargaining task, but not in other tasks. These beliefs turn out to be incorrect.

Table 6 estimates the relationship between an employer's performance expectation and worker attractiveness. Specifically, we regress the natural logarithm of the employer's prediction of worker performance in round 1 on the worker's beauty rating or on the indicator that the beauty rating is in a given quintile.

[TABLE 6 ABOUT HERE]

When we pool the data across tasks, we do not observe a significant relationship between worker attractiveness and employer performance prediction (Columns 1 and 2). As we anticipated, employers do not expect comparatively attractive workers to have a performance advantage in the data entry task (Columns 3 and 4). We also do not find a significant relationship between the worker's average attractiveness and expected performance in the data analysis task (Column 5). However, employers expect workers whose looks are in the 2nd quintile to perform significantly better than those in the bottom quintile (Column 6).[17] A positive relationship between beauty and expected performance emerges in bargaining (Column 7 and 8). A one standard deviation increase in attractiveness, on average, is associated with a 4% increase in the mean performance prediction. Similarly, in a linear specification, a one standard deviation increase in the beauty rating results in a statistically significant 2.8 point increase in predicted performance. The linear results can be found in the online appendix.

The effect appears to be driven by the high expectations for the most attractive workers: those in the 5th quintile are expected to perform 12% better in bargaining than the workers in the bottom quintile.[18] Once again, the linear specifications produce an even stronger positive relationship between beauty and predicted performance for the top attractiveness quintile (results available upon request). This finding is consistent with the "beauty-related" nature of the task, since workers can see their opponent's photos during bargaining.

Table 7 shows that employer expectations turn out to be incorrect in the bargaining task. We regress the natural logarithm of worker performance in round 1 on the average beauty rating or on the indicator that the beauty rating is in a given quintile. In this specification, standard errors are clustered by worker. Columns 7 and 8 (the bargaining task) also include a count variable for the number of bargaining periods during which trade was possible and control for the average difference between buyer value and seller cost across the three bargaining rounds. There is no systematic positive relationship between attractiveness and performance in any of the tasks, either on average or when we break the rating up into quintiles. Linear specifications with worker performance in levels produce qualitatively similar results. The results also do not change substantively if we omit the top attractiveness quintile instead of the bottom quintile.

[TABLE 7 ABOUT HERE]

[17] Employers also expect the second quintile to perform significantly better than the top quintile in data analysis (f-test p-value of 0.047).

[18] The results do not change if we omit the top quintile of attractiveness instead of the bottom quintile from the regressions.

Thus far, we have established that the beauty premium in bargaining is at least partly explained by employer beliefs about performance and that these beliefs are incorrect. We next proceed to test whether there is any taste-based discrimination by explicitly controlling for employer beliefs about performance in the round 1 wage bid (equation 1).

Result 3: The effect of beauty on wage bids disappears once we control for employer performance predictions, suggesting that there is no taste-based discrimination in our setting.

Support for Result 3 comes from Table 8, which estimates the effect of attractiveness on wage bids in round 1, controlling for the employer prediction of worker performance. Beauty is no longer a significant determinant of the wage bid in any of the tasks. Although the second and fourth quintiles are significant in the data entry task, the lack of a pattern in that regression suggests that the correlation is likely spurious. Our results are robust to controlling for polynomials of the employer performance prediction in addition to the performance prediction ranks.

[TABLE 8 ABOUT HERE]

As explained in Section 3, any residual relationship between beauty and wage bids can be interpreted as taste-based discrimination. Its absence in our setting suggests that employers are unwilling to sacrifice profits by hiring workers who are relatively attractive but not more productive. Overall, the evidence from Tables 5–8 shows that the statistical component of the beauty premium in the first round bargaining task can be explained by employers' biased beliefs about the performance of comparatively attractive workers, rather than tastes or rational statistical discrimination based on worker résumé characteristics.

4.2. Does Learning about Performance Eliminate the Beauty Premium in the Second Round?

Recall that Table 4 documents a significant relationship between attractiveness and wage bids in the second round, comparable in magnitude to that in the first round. So far, the evidence suggests that employers use appearance as a signal of ability, at least for the task that might be perceived as favoring comparatively attractive workers. However, we have also shown that the employers' beliefs are incorrect. Therefore, we proceed to examine the relationship between wage bids and beauty after relevant information about worker-specific previous performance is revealed to employers. Specifically, we estimate the effect of attractiveness on wage offers in the second round, with and without controlling for first-round performance information (see equation 1 in Section 3). We hypothesize that, because we don't observe a significant relationship between attractiveness and performance, we should observe a reduction in the beauty premium in the second round relative to the first, which would indicate learning.

Result 4: The beauty premium completely disappears in the second round. Information about past performance and employer expectations about future performance are both significant predictors of wage bids in the second round.

Support for Result 4 comes from Tables 9 and 10. Table 9 shows that, even when we do not control for past performance or the performance prediction (Columns 1, 3, 5, and 7), beauty is no longer correlated with wage bids in any of the tasks. This pattern holds as we add controls for performance in round 1 (Columns 2, 4, 6, and 8). Given our earlier findings that (a) employers use attractiveness as a signal of performance in the first round and (b) there is no relationship between attractiveness and performance, the absence of a beauty premium in the second round suggests that employers have learned that attractiveness is not a signal of ability in this setting and thus no longer utilize it as information from which to form wage bids. The specifications that break up the beauty rating into quintiles produce similar results. In particular, we find no evidence of the beauty premium for any of the quintiles. Similarly, restricting the sample to the sessions where the task in the second round was the same as in the first round does not change the results.

[TABLE 9 ABOUT HERE]

The estimated effects of performance in the first round on second round bids also confirm that there is substantial learning between rounds. Past performance is a significant predictor of bids in all tasks, with employers bidding more on workers who performed better in the first round, all else equal. This is true even in cases where the first round task is different from the second round task, implying that employers expect performance in the three tasks to be correlated.

Table 10 tests whether employer predictions of worker performance play a role in observed learning. Columns 1–4 show that employers no longer expect relatively more attractive workers to outperform less attractive workers in bargaining. As before, employers do not expect there to be a beauty advantage in the other two tasks. In fact, the point estimates on the average beauty rating are negative in the bargaining and data analysis task, although the negative value is not statistically significant.

Columns 5–8 of Table 10 build on the results from Table 9 to show the effect of employer expectations on wage bids in the second round. Once we control for employer performance predictions, past performance is significant only in the bargaining task, reaffirming that employers also fully incorporate first-round information into their performance expectations.[19] Thus, we conclude that learning eliminates the influence of beauty on the wage bid by changing employers' expectations of the relationship between beauty and worker performance.

[TABLE 10 ABOUT HERE]

4.3. Does the Beauty Premium Vary by Gender?

Table 4 shows that the simple correlation between attractiveness and wages varies by gender. In this section, we present a more formal analysis of the gender differences in the beauty premium.

Result 5: There is no difference in the beauty premium by worker gender. However, employers' performance expectations for comparatively attractive males and females differ from one another in data analysis and bargaining. These expectations turn out to be incorrect.

[19] Interacting first-round performance with an indicator that the second-round task was the same as the first-round task does not change the results.

Support for this result comes from Tables 11 and 12. Table 11 estimates the beauty premium in the first round by task and gender. When we do not control for performance expectations (Columns 1–3), we find a significant positive coefficient on female but not male attractiveness. However, the point estimates are very similar and the female and male beauty coefficients are not significantly different from one another. When we control for the rank of the performance prediction (Columns 4–6), there is no significant relationship between attractiveness and bids for either gender, suggesting that any relationship between attractiveness and bids is again driven by statistical discrimination.

[TABLE 11 ABOUT HERE]

Table 12 estimates the gender-specific relationship between employer performance expectations and attractiveness (Columns 1–3) and checks whether those expectations are correct (Columns 4–6). Although the beauty premium in bargaining does not vary by worker gender, employer expectations are different for comparatively attractive men than for comparatively attractive women. Relatively more attractive males are expected to perform marginally worse than less attractive males in the data analysis task, all else remaining equal. Furthermore, the expected beauty effects on data analysis performance for females (2.2%) and for males (-9.1%) differ significantly from one another (f-test p-value of 0.036). On the other hand, the positive effect of attractiveness of the employer prediction in bargaining from Table 6 is driven by the expectation that relatively more attractive males will have a performance advantage: comparatively attractive males are expected to outperform less attractive males by 10% in bargaining. Comparatively attractive females, on the other hand, are not expected to have a performance advantage in bargaining. The difference between males and females is statistically significant (f-test p-value of 0.001).

As before, the employer expectations turn out to be incorrect in data analysis and in bargaining: neither comparatively attractive males nor comparatively attractive females have an actual performance advantage in any of the tasks (Columns 4-6).

[TABLE 12 ABOUT HERE]

5. Conclusion

We develop and execute a new method for determining the precise channel through which attractiveness leads to higher worker wages. Our key insight is that having two measures, one that elicits expected worker performance and one that elicits employer willingness to pay, is both necessary and sufficient for separating statistical discrimination from taste-based discrimination without making restrictive assumptions. In addition, statistical discrimination can be further decomposed into biased beliefs and rational statistical discrimination if actual performance data are available.

We run a laboratory experiment designed to elicit the two measures described above as well as to identify the stability of the beauty premium across a range of settings. While carefully controlling the overall experimental environment, we vary the tasks that workers must perform. Our results indicate that the beauty premium is highly context-dependent: while we find strong evidence of a beauty premium in a bargaining task, there is no beauty premium in a data entry or data analysis task, on average. The beauty premium is composed entirely of statistical discrimination, which in turn can be explained by biased beliefs about the performance of comparatively attractive workers rather than rational expectations. It also does not appear to vary by gender: even though employers expect more attractive males to perform better in bargaining and worse in data analysis, these beliefs are incorrect. We also find a strong learning effect: the beauty premium disappears after worker performance is revealed, even in cases where the task changes. This suggests that, in our setting, employers use attractiveness at the hiring stage primarily as an imperfect signal of ability. Thus, one implication of our results is that the beauty premium in the labor market may be explained by worker characteristics and employer performance beliefs that cannot be fully incorporated into the analysis of labor market outcomes in observational data.

The absence of taste-based discrimination in our study may be explained in part by the minimal interactions between employers and workers in the experimental setting. Thus, our results may not generalize to situations in which there is substantial face-to-face contact: employers may be willing to pay more attractive workers higher wages due to taste on average, but do not do so in our case because they do not interact with workers in person. However, internet-based interactions are an increasingly important part of the modern economy. They are pervasive in online labor markets, such as oDesk, credit markets, such as Prosper, and even fundraising venues, such as Kickstarter. More generally, the spread of computers and the internet has transformed the modern workplace, with a growing fraction of workers spending all or most of their work time outside of the traditional office setting. Because of this trend, laboratory experiments, where subjects interact with each other largely through computers, are ever more relevant outside of the laboratory. Thus, our result pertaining to the absence of taste-based discrimination is highly applicable in these settings.

Because we do not find persistent biased beliefs in favor of more attractive people, the welfare losses from allowing beauty-based pay differentials are likely to be small in settings without substantial face-to-face interactions. Moreover, if the spectrum of real-world tasks exhibits beauty-based performance differentials (which we do not observe in our study), eliminating such differentials may lower the quality of matching between workers and jobs, leading to welfare losses. Testing for the existence of performance differentials across a number of jobs is an important step for future research.

Our understanding of the beauty premium may be enhanced by the introduction of face-to-face interactions between employers and workers and between the workers in the bargaining task in future experiments. Further exploration of the potential gift-exchange mechanism behind the beauty premium may also be a fruitful direction for future research.

References

Andreoni, J. and Petrie, R., 2008. Beauty, gender and stereotypes: evidence from laboratory experiments. *Journal of Economic Psychology*, vol. 29, 73–93.

Ayres, I. and Siegelman, P., 1995. Race and gender discrimination in bargaining for a new car, *American Economic Review*. 83(3), 304–321.

Becker, G., 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Berggren, N., Jordahl, H. and Poutvaara, P., 2010. The looks of a winner: beauty, gender, and electoral success. *Journal of Public Economics*, 94(1–2), 8–15.

Berri, D. J., Simmons, R., Van Gilder, J., and O'Neill, L., 2011. What does it mean to find the face of the franchise? Physical attractiveness and the evaluation of athletic performance. *Economic Letters*, 111, 200–202.

Biddle, J. and Hamermesh, D., 1998. Beauty, productivity, and discrimination: lawyers' looks and lucre. *Journal of Labor Economics,* vol. 16(1), 172–201.

Blanco, M., Engelmann, D., Koch, A. K., and Normann, H-T., 2010. Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, 13, 412–438.

von Bose, C., 2012. Child stars vs. ugly ducklings: does adolescent attractiveness contribute to the beauty premium? Working paper, The University of Texas at Austin.

Bottomley, L., Parry, E., Brigade, S., Coley, L., Deam, L., Goodson, E., Kidwell, J., Linck, J, and Robinson, B., 2001. Lessons learned from the implementation of a GK-12 grant outreach program. Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition, American Society for Engineering Education.

Castillo, M., Petrie, R., Torero, M., and Vesterlund, L., 2012. Gender differences in bargaining outcomes: a field experiment on discrimination. NBER Working paper No. 18093.

Chambers, D. W., 1983. Stereotypic images of the scientist: the draw the scientist test. *Science Education*, 67(2), 255–265.

Cooper, D. J. and Fang, H., 2008. Understanding overbidding in second price auctions: an experimental study. *The Economic Journal*, 118, 1572–1595.

Corbett, W. R., 2011. Hotness discrimination: appearance discrimination as a mirror for reflecting on the body of employment-discrimination law. *Catholic University Law Review*, 60 Cath. U.L. Rev. 615.

Deryugina, T. and Shurchkov, O., 2013. Is there a beauty premium in undergraduate education? Working paper, University of Illinois, Urbana-Champaign.

Duarte J., Siegel, S. and Young, L., 2012. Trust and credit: the role of appearance in peer-to-peer lending. *Review of Financial Studies,* 25(8), 2455–2484.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178.

Gächter, S. and Renner, E., 2010. The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13, 364–377.

Global Workplace Analytics, 2011. The state of telework in the US – five year trend and forecast. http://www.globalworkplaceanalytics.com/whitepapers

Hamermesh, D. and Biddle, J., 1994. Beauty and the labor market. *American Economic Review,* 84(5), 1174–1194.

Hamermesh, D., 2011. *Beauty Pays: Why Attractive People Are More Successful*. Princeton: Princeton University Press.

Mocan, N. and Tekin, E., 2010. Ugly criminals. *The Review of Economics and Statistics*, 92(1), 15–30.

Mobius, M. M. and Rosenblat, T. S., 2006. Why beauty matters. *American Economic Review*, 96(1), 222–235.

Pallais, A., 2013. Inefficient hiring in entry-level labor markets. Working paper, Harvard University.

Ravina, E., 2009. Love & loans: the effect of beauty and personal characteristics in credit markets. Working Paper, Columbia University.

Rosenblat, T., 2008. The beauty premium: physical attractiveness and gender in dictator games. *Negotiation Journal*, 24(4), 465–481.

Solnick, S. J., and Schweitzer, M. E., 1999. The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes*, 79(3), 199–215.

Wilson, R. K., and Eckel, C. C., 2006. Judging a book by its cover: beauty and expectations in a trust game. *Political Research Quarterly*, 59(2), 189–202.

Table 1
Treatment Summary

	Data Entry		Data Analysis		Bargaining	
	# Sessions	#Subjects	# Sessions	#Subjects	# Sessions	#Subjects
Round 1	16	128	15	120	16	128
Round 2	15	120	16	128	16	128
Round 1 = 2	7	56	8	64	8	64

Notes: Round 1 and Round 2 rows list all sessions, whether the task was the same or different in the second round. Sessions with the same task in both rounds are listed in the row labeled "Round 1 = 2."

Table 2
Summary statistics of experimental outcomes by task and gender

	Data Entry		Data Analysis		Bargaining	
	Male	Female	Male	Female	Male	Female
	Panel 1: Averages for Round 1					
Employers						
Wage Bid (Points)	27.6	27.9	23.6	24.6	27.9	28.9
Average Prediction (Points)	70.1	73.3	10.4	11.4	61.6	64.2
Payoff (Points)	89.4	91.4	98.8	87.6	77.6	80.5
Observations	18	41	19	38	22	36
Workers						
Own Prediction (Points)	88.6	89.4	10.5	9.5	68.7*	60.4*
Performance (Points)	83.6	91.5	7.2	7.2	79.5**	55.5**
Payoff (Points)	108.1	107.9	72.7	78.6	102.5**	83.2**
Observations	15	44	36	23	26	33
	Panel 2: Averages for Round 2					
Employers						
Wage Bid (Points)	32.5	30.3	22.1	23.7	25.5*	30.2*
Average Prediction (Points)	85.9**	78.3**	9.2	9.1	70.4	64.2
Payoff (Points)	91.2	90.7	88.2	93.2	78.7	68.4
Observations	21	37	20	39	18	39
Workers						
Own Prediction (Points)	106.0	92.8	10.2	9.8	68.3	66.1
Performance (Points)	86.3	84.9	11.1	8.5	70.3	63.7
Payoff (Points)	103.8	116.4	83.9*	72.0*	92.7	81.9
Observations	22	36	36	24	19	40

Notes: The means for Employers are separated by male and female employers. The means for Workers are separated by male and female workers. Significance levels based on t-tests of differences between male and female subjects: * 10 percent, ** 5 percent. A Mann-Whitney U test that compares distributions produces similar results.

Table 3

Summary statistics of attractiveness and other subject characteristics

	Employers		Workers	
	Male	Female	Male	Female
Demeaned Attractiveness Rating	-0.011	0.010	-0.067	0.054
	Panel 1: Resume Characteristics			
Analytical Skills	1.76***	1.56***	1.66	1.59
Typing Skills	1.40***	1.66***	1.40	1.53
Communication Skills	1.76*	1.65*	1.60	1.61
Resume GPA	2.39**	2.58**	2.49	2.57
Observations	59	115	77	100
	Panel 2: Other Characteristics			
Exact GPA	3.45**	3.54**	3.51	3.54
Major	2.22	2.15	2.17	2.14
Share Native English Speakers	0.93**	0.79**	0.86	0.83

Notes: Skills are measured on a scale of 0–2, with 2 representing excellent. Resume GPA is measured on a scale of 0–3, with 3 representing the range 3.5–4.0. Major is 1 for Humanities; 2 for Social Sciences; 3 for Natural Sciences. Significance levels based on t-tests of differences between male and female subjects: * 10 percent, ** 5 percent, ***1 percent.

Table 4

The beauty premium by task for all workers and by worker gender in both rounds

Outcome variable:	Natural logarithm of employer wage bid							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks	Data Entry	Data Analysis	Barg.	All Tasks	Data Entry	Data Analysis	Barg.
	ROUND 1				ROUND 2			
	Panel 1: All workers							
Attractiveness of worker	0.162***	0.087	0.144	0.235**	0.158***	0.176	0.086	0.231**
	(0.056)	(0.093)	(0.106)	(0.094)	(0.056)	(0.132)	(0.082)	(0.091)
Observations	685	232	225	228	685	225	236	224
R-squared	0.04	0.03	0.05	0.05	0.04	0.03	0.05	0.05
	Panel 2: Males versus females							
Attractiveness of worker								
if female	0.197***	0.041	0.261**	0.328***	0.229***	0.524***	0.103	0.160
	(0.065)	(0.103)	(0.126)	(0.105)	(0.079)	(0.160)	(0.125)	(0.126)
if male	0.116	0.306	0.037	0.211	0.052	-0.384	0.066	0.422*
	(0.096)	(0.265)	(0.151)	(0.143)	(0.096)	(0.254)	(0.091)	(0.227)
F-test p-value (equality)	[0.470]	[0.367]	[0.237]	[0.473]	[0.188]	[0.004]	[0.804]	[0.373]
Observations	685	232	225	228	685	225	236	224
R-squared	0.04	0.03	0.06	0.08	0.04	0.09	0.05	0.05

Notes: The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. All regressions include date fixed effects. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 5

Relationship between an employer bid in round 1 and worker attractiveness

Outcome variable:	Natural logarithm of employer wage bid in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Entry		Data Analysis		Bargaining	
	[26.9]		[27.9]		[24.3]		[28.5]	
Attractiveness of worker	0.136**		0.011		0.159		0.280**	
	(0.060)		(0.114)		(0.130)		(0.112)	
Attractiveness quintiles:								
2nd		0.026		0.752*		-0.469		0.384
		(0.205)		(0.381)		(0.382)		(0.487)
3rd		0.101		0.671*		0.409		0.357
		(0.187)		(0.384)		(0.371)		(0.450)
4th		0.395**		0.974**		0.512		0.572
		(0.182)		(0.393)		(0.330)		(0.432)
Top attractiveness: 5th		0.305		0.16		-0.022		0.996**
		(0.212)		(0.414)		(0.352)		(0.435)
F-test p-value (equality)		[0.136]		[0.171]		[0.024]		[0.248]
Observations	685	685	232	232	225	225	228	228
R-squared	0.16	0.17	0.19	0.22	0.26	0.29	0.20	0.21

Notes: Round 1 data only. Mean wage bids (in points) for each task are reported in brackets below the task type. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. The p-values for the f-tests of joint differences of the coefficients on the attractiveness quintiles are reported in brackets below the estimates. All regressions include date fixed effects, indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 6

Relationship between employer performance expectations and worker attractiveness in round 1

Outcome variable:	Natural logarithm of employer performance prediction in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Entry		Data Analysis		Bargaining	
	[63.6]		[72.2]		[55.5]		[63.0]	
Attractiveness of worker	0.022		-0.016		-0.192		0.043*	
	(0.018)		(0.041)		(0.151)		(0.025)	
Attractiveness quintiles:								
2nd		-0.020		-0.037		0.160**		-0.132
		(0.077)		(0.054)		(0.080)		(0.253)
3rd		-0.068		-0.074		0.010		-0.136
		(0.057)		(0.149)		(0.064)		(0.170)
4th		0.030		-0.074		0.081		0.110
		(0.052)		(0.086)		(0.061)		(0.131)
Top attractiveness: 5th		0.050		-0.035		-0.025		0.118*
		(0.042)		(0.069)		(0.074)		(0.072)
F-test p-value (equality)		[0.360]		[0.950]		[0.126]		[0.208]
Observations	685	685	232	232	225	225	228	228
R-squared	0.22	0.23	0.33	0.33	0.28	0.29	0.22	0.23

Notes: Round 1 data only. Mean predicted performance (in points) for each task is reported in brackets below the task type. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. The p-values for the f-tests of joint differences of the coefficients on the attractiveness quintiles are reported in brackets below the estimates. All regressions include date fixed effects, indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 7
Relationship between a worker attractiveness and performance in round 1

Outcome variable:	Natural logarithm of worker performance in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Entry		Data Analysis		Bargaining	
	[63.8]		[89.5]		[35.8]		[66.1]	
Attractiveness of worker	0.004		-0.121		-0.022		-0.054	
	(0.047)		(0.083)		(0.051)		(0.167)	
Attractiveness quintiles:								
2nd		0.028		-0.143		0.122		-0.612
		(0.166)		(0.326)		(0.184)		(0.728)
3rd		-0.114		-0.352		-0.028		-0.256
		(0.198)		(0.294)		(0.164)		(0.613)
4th		0.152		-0.319		-0.05		0.459
		(0.163)		(0.373)		(0.159)		(0.717)
Top attractiveness: 5th		-0.038		-0.507		-0.015		-0.25
		(0.174)		(0.329)		(0.212)		(0.593)
F-test p-value (equality)		[0.399]		[0.154]		[0.582]		[0.290]
Observations	177	177	59	59	59	59	59	59
R-squared	0.39	0.40	0.54	0.59	0.68	0.69	0.50	0.55

Notes: Round 1 data only. Mean worker performance (in points) for each task is reported in brackets below the task type. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. The p-values for the f-tests of joint differences on the coefficients of attractiveness quintiles are reported in brackets below the estimates. All regressions include date fixed effects, indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. Specifications (7) and (8) include an indicator for whether a trade was possible and control for the average difference between buyer value and seller cost across the three bargaining rounds. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by worker in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 8

Separating statistical from taste-based discrimination

Outcome variable:	Natural logarithm of employer wage bid in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Entry		Data Analysis		Bargaining	
Attractiveness of worker	0.081		-0.018		0.136		0.078	
	(0.056)		(0.105)		(0.127)		(0.113)	
Attractiveness quintiles:								
2nd		0.12		0.775*		-0.474		0.504
		(0.192)		(0.399)		(0.349)		(0.451)
3rd		0.085		0.526		0.43		0.336
		(0.172)		(0.405)		(0.353)		(0.380)
4th		0.379**		0.869**		0.413		0.34
		(0.166)		(0.388)		(0.309)		(0.412)
Top attractiveness: 5th		0.225		0.292		-0.039		0.373
		(0.184)		(0.413)		(0.320)		(0.387)
F-test p-value (equality)		[0.238]		[0.174]		[0.053]		[0.980]
Employer's performance								
prediction rank: 2nd	0.652***	0.652***	0.842***	0.826***	0.231	0.181	0.799**	0.834***
	(0.164)	(0.164)	(0.268)	(0.270)	(0.290)	(0.293)	(0.306)	(0.313)
3rd	1.368***	1.374***	1.162***	1.171***	0.820***	0.749***	1.849***	1.867***
	(0.172)	(0.170)	(0.282)	(0.280)	(0.2780)	(0.269)	(0.298)	(0.293)
Top prediction rank: 4th	1.680***	1.685***	1.538***	1.478***	1.279***	1.230***	2.010***	2.059***
	(0.187)	(0.186)	(0.317)	(0.313)	(0.356)	(0.352)	(0.342)	(0.341)
F-test p-value (joint sign.)	[0.0000]	[0.0000]	[0.0001]	[0.0002]	[0.0054]	[0.0065]	[0.0000]	[0.0000]
Observations	685	685	232	232	225	225	228	228
R-squared	0.30	0.31	0.31	0.33	0.32	0.35	0.39	0.40

Notes: Round 1 data only. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. All regressions include date fixed effects, indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 9

Relationship between an employer bid in round 2 and worker attractiveness

Outcome variable:	Natural logarithm of employer wage bid in round 2							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Entry		Data Analysis		Bargaining	
Attractiveness of worker	0.100	0.102*	0.285	0.319	0.213	0.107	0.028	0.065
	(0.063)	(0.061)	(0.241)	(0.234)	(0.146)	(0.122)	(0.132)	(0.133)
Log performance in round 1		0.815***		0.808**		1.058***		0.664***
		(0.085)		(0.338)		(0.180)		(0.136)
Observations	685	685	225	225	236	236	224	224
R-squared	0.09	0.20	0.20	0.23	0.25	0.35	0.19	0.28

Notes: Round 2 data only. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. All regressions include date fixed effects, indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 10

The role of employer performance expectations in the second round

Outcome variable:	Natural logarithm of employer prediction in round 2				Natural logarithm of employer wage bid in round 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks	Data Entry	Data Analysis	Barg.	All Tasks	Data Entry	Data Analysis	Barg.
Attractiveness of worker	-0.010	0.012	-0.011	-0.061	0.087	0.146	0.100	0.033
	(0.018)	(0.045)	(0.032)	(0.043)	(0.056)	(0.177)	(0.092)	(0.105)
Log performance in round 1	0.460***	0.255**	0.396***	0.519***	0.304***	0.121	0.450**	0.357**
	(0.066)	(0.107)	(0.063)	(0.109)	(0.087)	(0.290)	(0.177)	(0.155)
Employer's performance prediction rank: 2nd					0.925***	1.324***	0.853***	0.700**
					(0.166)	(0.318)	(0.286)	(0.308)
3rd					1.581***	2.023***	1.570***	1.026**
					(0.182)	(0.457)	(0.236)	(0.412)
Top prediction rank: 4th					2.109***	2.238***	1.982***	1.675***
					(0.183)	(0.391)	(0.315)	(0.385)
Observations	685	225	236	224	685	225	236	224
R-squared	0.50	0.33	0.44	0.50	0.36	0.37	0.49	0.34

Notes: Round 2 data only. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. All regressions include date fixed effects, indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. In cases where we pool the data across multiple tasks, we also include a task fixed effect. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 11

Relationship between employer bid in round 1 and worker attractiveness by worker gender

Outcome variable:	Natural logarithm of employer wage bid in round 1					
	(1)	(2)	(3)	(4)	(5)	(6)
	Data Entry	Data Analysis	Barg.	Data Entry	Data Analysis	Barg.
Attractiveness of worker						
if female	-0.031	0.170	0.285**	-0.065	0.155	0.141
	(0.137)	(0.150)	(0.123)	(0.127)	(0.144)	(0.124)
if male	0.169	0.150	0.274	0.154	0.119	-0.009
	(0.286)	(0.177)	(0.201)	(0.264)	(0.175)	(0.195)
F-test p-value (equality)	[0.557]	[0.923]	[0.961]	[0.490]	[0.856]	[0.522]
Employer's performance						
prediction rank: 2nd				0.844***	0.232	0.797**
				(0.267)	(0.291)	(0.303)
3rd				1.175***	0.823***	1.855***
				(0.278)	(0.280)	(0.298)
Top prediction rank: 4th				1.529***	1.277***	2.033***
				(0.316)	(0.358)	(0.336)
Observations	232	225	228	232	225	228
R-squared	0.20	0.26	0.20	0.31	0.32	0.39

Notes: Round 1 data only. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. The p-values for the f-tests of joint differences of the coefficients on the gender interactions are reported in brackets below the estimates. All regressions include date fixed effects and indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. Robust standard errors are clustered by employer in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 12

Determinants of employer expectations about worker performance and actual performance by gender

Outcome variable:	Natural log of employer performance prediction in round 1			Natural log of actual worker performance in round 1		
	(1)	(2)	(3)	(4)	(5)	(6)
	Data Entry	Data Analysis	Barg.	Data Entry	Data Analysis	Barg.
Attractiveness of worker						
if female	0.003	0.022	-0.005	-0.090	0.033	-0.193
	(0.041)	(0.023)	(0.020)	(0.087)	(0.076)	(0.191)
if male	-0.087	-0.091*	0.105***	-0.243	-0.069	0.155
	(0.068)	(0.048)	(0.035)	(0.154)	(0.069)	(0.232)
F-test p-value (equality)	[0.152]	[0.036]	[0.001]	[0.370]	[0.329]	[0.208]
Observations	232	225	228	59	59	59
R-squared	0.33	0.29	0.22	0.55	0.68	0.51

Notes: Round 1 data only. The attractiveness coefficient should be interpreted as the effect of a one standard deviation change in beauty on the outcome variable. The p-values for the f-tests of joint differences of the coefficients on the gender interactions are reported in brackets below the estimates. All regressions include date fixed effects and indicators for student status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. Column (6) includes an indicator for whether a trade was possible and control for the average difference between buyer value and seller cost across the three bargaining rounds. Robust standard errors are clustered by employer (Columns 1–3) or worker (Columns 4–6) in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.