Now you see it, now you don't: the vanishing beauty premium[*]

Tatyana Deryugina, University of Illinois[+]

Olga Shurchkov, Wellesley College[++]

First Draft: December 2012

Current Draft: April 2015

Abstract:

We design a laboratory experiment to test the extent to which the often-observed "beauty premium" – a positive relationship between attractiveness and wages – is context-specific. Using three realistic worker tasks, we find that the existence of the "beauty premium" indeed depends on the task: 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. Our analysis shows that the premium in bargaining is driven by statistical discrimination based on biased beliefs about worker performance. We also 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

---

[*] We are grateful to Kristin Butcher, Phillip Levine, Robin McKnight, Julian Reif, Casey Rothschild, Gauri Kartini Shastri, Akila Weerapana and participants in the Wellesley Economics Department work-in-progress seminar and the University of Illinois at Urbana-Champaign Finance seminar for valuable comments and discussion. In addition, we would like to acknowledge the generous research support we received from the Harvard Decision Science Laboratory to organize and run our experiment. Shurchkov would like to acknowledge a Harvard Kennedy School's Women and Public Policy Program fellowship for providing the resources that made this research possible. Lizi Chen, Divya Gopinath, Mehrnoush Shahhosseini, and Sizhe Zhang provided excellent research assistance. All remaining errors are our own. This paper was previously circulated under the title "When are Appearances Deceiving? The Nature and Evolution of the Beauty Premium."
[+] Department of Finance, University of Illinois at Urbana-Champaign, 1206 South Sixth St., Champaign, IL, USA.
e-mail: deryugin@illinois.edu.
[++] Department of Economics, Wellesley College, 106 Central St., Wellesley, MA, USA.
e-mail: olga.shurchkov@wellesley.edu.

1. Introduction

Labor market discrimination based on gender, age, race, and national origin is illegal. Appearance-based discrimination, while not currently unlawful, has been the subject of several lawsuits in recent years.[1] Supporting the notion that appearance-based discrimination exists, numerous observational studies have found that people who are relatively more attractive are paid more, even when the situation does not appear to warrant it. This phenomenon has been termed the "beauty premium." It appears to be pervasive: versions of the beauty premium have been found in labor markets (e.g., Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998), college classrooms (Hamermesh and Parker, 2005; Sen et al., 2010; Ponzo and Scoppa, 2012), credit markets (Ravina, 2012), sex markets (Arunachalam and Shah, 2012), professional sports (Berri et al., 2011), and elections (Hamermesh, 2006; Leigh and Susilo, 2009; Berggren et al., 2010).

One potential explanation for the beauty premium in naturally occurring data is that appearance may in fact be positively correlated with skills that are important for job performance but are not easily observed, such as the ability to be persuasive ("statistical discrimination"). Another is that employers may have biased beliefs, overestimating the skills of relatively attractive people. Finally, employers may have unbiased beliefs about performance but prefer hiring more attractive people ("taste-based discrimination").

We use a novel approach to separate taste-based discrimination from statistical discrimination and biased beliefs in a laboratory labor market. First, we directly elicit beliefs about each worker's performance, which allows us to determine what share of the beauty premium, as measured by employers' wage bids on workers, is statistical discrimination. Then, by controlling for performance predictions, we are able to estimate the portion of the wage bid that is *not* driven by performance expectations and test whether it is correlated with the worker's attractiveness. Finally, because we observe workers' actual performance, we can 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 relatively attractive people.

Another innovation of our study is to estimate the size of the beauty premium across three different labor-market relevant tasks: a data entry task, an analytical task, and a bargaining task in which workers see pictures of their bargaining opponents. Our study is the first to explicitly test whether the beauty premium varies with the types of skills involved in completing a task and, if so, to determine why.

To our knowledge, we are also the first to examine learning in the context of the beauty premium in a labor market.[2] It is possible that attractiveness is used as a proxy for ability when job-specific information about a worker's performance is scarce. We model such scarcity in our experimental setting with the first round, where employers only observe resumes and photos. However, attractiveness could become increasingly irrelevant as employers learn about 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.

[1] See for example Yanowitz v. L'Oreal USA, Inc. (2005) and Brice v. Resch and Krueger Int'l, Inc. (Corbett, 2011).

[2] See Wilson and Eckel (2006) for beauty and learning in a trust game and Andreoni and Petrie (2008) and Castillo et al. (2012a) for beauty and learning in a public goods game, among others.

Our analysis yields three key findings. First, there is a significant beauty premium in bargaining but not in data analysis or data entry. In particular, a one-standard-deviation increase in worker attractiveness is associated with a 26.5 percent 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 command the highest beauty premium in bargaining. On the other hand, the most attractive workers suffer a beauty penalty in data entry.[3] Our conclusion that the beauty premium is highly context-specific is consistent with some 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, 2013; Deryugina and Shurchkov, 2015).

Second, we find that 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 a worker's attractiveness and performance in any of the tasks.

Finally, we find that the beauty premium in bargaining completely vanishes in the second round of bidding when the task is repeated, 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 because it affects employer beliefs about future 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. Our results are consistent with previous evidence that discrimination based on individual characteristics is more likely to occur in the absence of information. For example, Castillo and Petrie (2010) study group formation in a public goods game experiment and find that information about behavior causes people to disregard personal characteristics such as race and appearance. In another study, Berggren et al. (2010) use data on outcomes in Finnish parliamentary elections and find that the beauty premium exists only for non-incumbents, which implies that availability of performance data for the incumbents eliminates the effect of attractiveness.

Our laboratory study complements existing literature that uses observational data to study the beauty premium in labor markets (e.g., Hamermesh and Biddle, 1994; Biddle and Hamermesh, 1998; Harper, 2000; Fletcher, 2009; Hamermesh, 2011; Borland and Leigh, 2014; and Scholz and Sicinski, 2015). Although most studies find a positive effect of beauty on earnings, Harper (2000) finds that the earnings advantage of attractive individuals disappears once academic ability and sociability are controlled for. However, the earnings penalty of relatively unattractive people persists. While Pope and Sydnor (2011) do not find an effect of looks in an online credit market, Ravina (2012) uses a finer measure of attractiveness in the same setting and finds a significant beauty premium. The advantage of an experimental setting is that we can vary the environment to determine the conditions that would lead to the existence of the beauty premium.

Our experiment is most closely related to Mobius and Rosenblat (2006), hereafter MR 2006, who also examine the nature of the beauty premium in an experimental labor market where "employers" determine the wages of "workers." However, our study differs from MR 2006 in several important ways. First, we vary the task the workers are asked to perform, allowing for the possibility that the relationship between attractiveness and performance varies by task. Second,

[3] Dermer and Thiel (1975) document that certain socially undesirable characteristics, such as egotism and materialism, are ascribed to relatively attractive females. Ruffle and Shtudiner (2014) find a beauty penalty for attractive female job applicants. Wilson and Eckel (2006) find evidence of a beauty penalty in trust games.

MR 2006 use variation in whether the employer's performance prediction determines the worker's wage to separate taste-based from statistical discrimination and do not elicit wage bids separately. By contrast, our identification of the two types of discrimination is fundamentally different, as we elicit both wage bids and performance predictions. In addition, employers are competing to hire workers in our setting. Finally, we observe employers' performance predictions and wage bids both before and after they observe workers' performance. However, we do not vary the means by which employers and workers interact (e.g., face-to-face versus over the phone): in our setting, employers never talk to or see the workers in person.

Our general approach shows that having measures of both (a) expectations about a worker's performance and (b) willingness to hire that worker (by giving her a relatively higher wage offer) are helpful for separating different kinds of discrimination. Prior research has used various methods for distinguishing taste-based from statistical discrimination that do not involve eliciting beliefs directly. For example, in a field study on gender differences in bargaining outcomes over taxi fares in Peru, Castillo et al. (2012b) infer beliefs from observed initial price quotes. In a quasi-experiment using a popular British game show, Belot et al. (2012) infer that discrimination against homely contestants must not be statistical because performance is not related to attractiveness. Laboratory studies that distinguish taste-based from statistical discrimination without eliciting beliefs include Fershtman and Gneezy (2001), Castillo and Petrie (2010), and Castillo et al. (2012a).

Direct elicitation of beliefs may be preferable for three reasons. First, and most importantly, we do not have to assume that the learning process is perfect to estimate the role that beliefs play in behavior. Because people have been shown to violate Bayes' rule in numerous situations, belief elicitation allows the researcher to test directly how much learning has occurred.[4] Eliciting beliefs may be particularly helpful when it is infeasible to provide subjects with many (or perhaps any) learning opportunities, as in our case. Second, eliciting beliefs allows the researcher to study them directly, and determine whether subjects hold correct or biased beliefs. Third, controlling for beliefs can help determine whether beliefs fully explain behavior or whether taste-based discrimination is also present.

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, 2014), and online credit markets, such as Prosper (Ravina, 2012; Duarte et al., 2012). Moreover, with the increasing number of "telecommuting" workers, the old paradigm of "face time" is changing. 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, 2014). In the US, telecommuting increased 73% from 2005 to 2011, and 64 million U.S employees holds a job that is compatible with at least part-time telework (Global Workplace Analytics, 2011). In addition, hiring in some cases is done by temp agencies or human resources departments, which are functionally removed from daily interactions with the

[4] For an overview of non-Bayesian updating, see Kahneman et al. (1982) and Rabin (1998).

hired workers.[5] Our findings about the absence of 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. We ran a total of 45 experimental sessions. Each session included four employers and four workers. In total, 180 employers and 180 workers participated in our experiment. We drop two workers who chose to withdraw from our study after the experiment. The final dataset consists of 180 employers and 178 workers.

The first four subjects to arrive at the laboratory ("employers") were immediately taken from the waiting room, photographed, and seated at their stations. The next four subjects to arrive ("workers") were photographed and seated afterwards. 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. Most subjects arrived in the laboratory within ten minutes of one another, which ensured plausibly random role assignment. 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.

After being seated, all subjects answered survey questions about several characteristics that are relevant to the labor market (student status, major, GPA range, as well as levels of typing, analytical, and communication skills) before being told whether they would be employers or workers. Summary statistics of various employer and worker characteristics can be found in the online appendix. After receiving the experimental instructions, which included detailed information about the task workers would perform, employers were granted access to a website that displayed worker photographs and the corresponding "résumés" built from each worker's survey answers. In 25 sessions, photos were shown first, with links to résumé information underneath each photo. In 22 sessions, this order was reversed.

The major difference between sessions was in the tasks workers had to perform, which were randomly assigned for each session. During the experimental design stage, we determined the number of different task combinations that workers would perform (e.g., bargaining-bargaining, data entry-bargaining, data analysis-data entry, etc.). To ensure that the sessions were balanced, the overall choice of task combinations was not random. The goal was to have each task be performed approximately the same number of times overall, in combination with other tasks, and across rounds. (For more details, see Section 2.2 below).

[5] Ruffle and Shtudiner (2014) and López Bóo et al. (2013) investigate the role of physical attractiveness at this initial hiring stage by examining the response rates to fictional CVs. Ruffle and Shtudiner (2014) find a gender-specific effect of attractiveness, namely a beauty premium for males and a beauty penalty for females, while López Bóo et al. (2013) find a beauty premium for both genders. The beauty penalty observed by Ruffle and Shtudiner (2014) is most consistent with jealousy, a type of taste-based discrimination that may not arise in our setting due to limited interactions between workers and employers. It is unclear whether the beauty premium found by López Bóo et al. (2013) is based on tastes or is statistical.

The remainder of the experiment, programmed using the standard zTree software package (Fischbacher, 2007), consisted of two procedurally identical rounds. First, instructions for the task the workers would perform in that round were distributed and read. Neither the workers nor the employers were told what task would be performed in the second round until after the first round was over. Then, employers submitted estimates for the expected performance of each worker in the subsequent task (E_{ij}), where i indexes employers and j indexes workers. At the same time, workers submitted estimates for their own expected performance (E_j). This information was kept secret from all other subjects. The payoffs of both employers and workers were partly determined by the accuracy of their predictions, ensuring that they had incentives to guess correctly.

Next, employers submitted wage bids to "hire" workers. The total amount offered to four workers could not exceed a predetermined maximum number of points. In the first 22 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. We find no evidence that this change affected bidding behavior, however.

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 was not revealed to the worker until after the task completion stage. By withholding the wage offer information until after the worker completes the task, we are shutting down any "gift-exchange" mechanism behind the beauty premium, i.e., employers expecting more attractive workers to reciprocate a higher wage with higher effort. 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 in trust games has found that more attractive individuals do not exhibit greater levels of reciprocity relative to their less attractive counterparts (Wilson and Eckel, 2006).

If employers are rational, the sequential elicitation of performance predictions and wage bids should not affect the way in which employers make either decision. If behavioral factors are at play, elicitation of performance predictions prior to the submission of wage offers may increase the salience of productivity expectations and subsequently cause these expectations to play a larger role in the employers' wage bid decision. This would bias us toward *not* finding taste-based discrimination relative to an alternative design where performance expectations are not elicited explicitly. Although testing for such an effect is outside the scope of this paper, the possibility that simply making performance expectations salient would reduce taste-based discrimination is an intriguing hypothesis for future research.

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, although neither employers nor workers were aware of their matches or wage payments at this point.

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:

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

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; Y_j is the output, in points, of worker *j*; 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 (MR 2006). Incentivized belief elicitation may distort worker incentives during the task completion stage, leading to a "hedging bias" (Blanco et al., 2010). 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.

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

In each round of a session, workers had to perform one the following three tasks: data analysis, data entry, or bargaining. Across all sessions, each task was performed exactly 15 times in the first round and 15 times in the second round. We had 3-4 sessions of each possible combination of *different* tasks (e.g., data entry-data analysis, bargaining-data entry) and 7-8 sessions of each combination of the same task (bargaining-bargaining, data entry-data entry, data analysis-data analysis). The assignment of each task combination to experimental sessions was random, resulting in as-good-as-random assignment of the first- and second-round tasks from the subjects' point of view. Table 1 shows the number of sessions for each task type and the corresponding number of subjects who participated in a given session. For an exact breakdown of task combinations, see the online appendix.

[TABLE 1 ABOUT HERE]

In order to isolate the impact of the nature of the task on the beauty premium, we strove to equalize the amount of time the tasks took to perform as well as the expected earnings of employers and workers. Due to the relative difficulty of the data analysis task, we allowed subjects twice as much time to complete it as the other tasks. Pilot sessions were used to gauge average performance in each task and calibrate payoffs to equalize earnings. However, as we later discuss, payoffs did vary slightly across tasks at the conclusion of the experiment.

We deliberately chose task types with which employers are more likely to be familiar and, thus, in which appearance-based differences in expectations are more likely to be correct. Having workers perform realistic tasks also increases the external relevance of our study.

In the data entry task, workers had six minutes to enter numerical data from a sheet of paper into an Excel spreadsheet. The data consisted of regional economic statistics for Russia. The spreadsheets had already been opened on the workers' computers and the column and row headings had been prepared in advance, so that subjects only had to enter numerical values into the correct cells. The data had to be entered exactly as it appeared to receive credit. Workers received one point per correctly entered item, creating an incentive to enter as much data as possible. There was no penalty for an incorrectly entered item.

In the data analysis task, workers answered up to 30 mathematical questions. Questions were based on data 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, simple calculators were provided. Workers received five points per correctly answered question, and there was no penalty for answering questions incorrectly.

In the bargaining task, workers were randomly assigned to be 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 for each bargaining round, which was not part of the 90-second limit, the bargaining task also lasted about six minutes. Workers were randomly re-matched and roles were randomly assigned with every new bargaining period.

During bargaining, each worker saw a photo of his or her bargaining partner on a computer screen. This feature distinguishes bargaining from the other two tasks in that appearance may be a direct factor in performance. 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.

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. Buyers' values and sellers' costs were determined randomly from two uniform distributions. In some cases, the buyer's value was below the 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.

The questions asked in data analysis and the data used in data entry did not vary across sessions or rounds, with the exception of sessions that had data entry or data analysis in both rounds. In these cases, we used a different set of questions/data to be entered to avoid making the second-round task too easy or too repetitive. Due to the random assignment of seller and buyer values in each instance of bargaining, it was not necessary to alter the bargaining task.

2.3. Rating Procedures

Attractiveness of the experimental subjects was evaluated by students ("raters") at the University of Illinois, Urbana-Champaign. During each session, 4–15 raters were instructed to evaluate photos on a scale from 1 (homely) to 10 (strikingly handsome or beautiful). Each rater was asked to look through four sets of 100 photos from this and another beauty-related study (Deryugina and Shurchkov, 2015), 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. Summary statistics of rater demographics can be found in the online appendix.

2.4. Descriptive Statistics

If employers behave similarly and worker performance is on average the same across tasks, employer earnings should be roughly equal to each other in all cases, as should worker earnings. Three effects could lead to differences in earnings across tasks. First, worker performance differences across tasks would translate into earnings differences for both workers and employers, all else equal. Second, employers' bidding strategies could vary across tasks. For example, if employers bid unjustifiably more on more attractive workers in bargaining, this could adversely affect their earnings while increasing the earnings of workers. Finally, if employers or workers make systematically biased predictions in one of the tasks, the prediction penalty would lead to lower earnings in that task.

In round 1, the average employer wage bid in data analysis is statistically lower than the wage bid in data entry and bargaining. The pattern is similar in round 2. In both rounds, employers predict lower performance in data analysis (in points) relative to the other two tasks. Actual worker performance in data analysis in both rounds is indeed significantly lower, on average. The relevant summary statistics can be found in the online appendix.

Although we aimed to make payoffs comparable across tasks, differences arose. In both rounds, workers earned significantly lower payoffs in data analysis relative to the other two tasks, while employers earned significantly lower payoffs in bargaining. In order to remove any potential bias stemming from these task differences, we estimate within-task effects of beauty by including task fixed effects in pooled analysis or by looking at each task separately.

Before proceeding with formal regression analysis, we test for the existence of a relationship between wage bids and attractiveness without controlling for worker characteristics. Table 2 reports the results for both rounds.

[TABLE 2 ABOUT HERE]

When pooling all tasks (Columns 1 and 5), we find a significant positive relationship between the natural logarithm of the wage bid and worker attractiveness in both rounds. On average, a one standard deviation increase in attractiveness increases the wage bid by about 17 percent in round 1 and 15 percent in round 2. When we decompose this relationship by task, we do not find any statistically significant effect of beauty on 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 24 percent in round 1 and 21 percent in round 2. Later in the paper, we elucidate the mechanisms behind this task-specific relationship.

3. Empirical Strategy

In this section, we outline our empirical strategy to explore the mechanisms behind the beauty premium we found in Table 2. In particular, we aim to (1) separate statistical discrimination from taste-based discrimination; (2) test whether statistical discrimination, if any, is based on rational or biased beliefs; and (3) gauge the extent of learning by employers. Our main unit of observation for studying bidding behavior is employer-worker pairs, of which we have 16 per session (four employers each bidding on four workers), for a total of 712 pairs.

To separate taste-based 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, in the case of purely statistical discrimination, once we properly control for the performance prediction, worker's attractiveness should have no further explanatory power. 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: employers bid more on more attractive workers even though they do not expect them to be more productive.

Unfortunately, the optimal bidding strategy in our setting is not analytically tractable. Moreover, prior experimental literature finds that behavior consistently deviates from rational bidding strategies (e.g., Cooper and Fang, 2008). Because an employer's bid for a worker may well be nonlinear in her performance prediction, we allow the performance prediction to enter the specification flexibly, as the employer's rank of the worker's performance, ranging from 1 to 4.[6] Specifically, the regression specification is as follows:

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

where i indexes the employer and j indexes the worker. We suppress the round subscripts for tractability. The variable W_{ij} is the bid of employer i on worker j in round 1 or round 2. The natural logarithm specification addresses skewness in the wage bids. Wages are typically right skewed, both in observational data and in our experimental setting. Taking the natural logarithm avoids having estimates being driven by outliers (see, for example, Amiti and Davis, 2012; Autor et al., 2013; Walker, 2013; and Card et al., 2014). Specifications that use levels do not produce substantively different results and are available in the online appendix.

The attractiveness rating is given by a_j, and $E_i[P_j]$ is employer i's expectation of worker j's performance. Worker resume characteristics are captured by X_j' and include indicators for student

[6] Using a third degree polynomial instead of rank to control for the performance prediction produces similar results.

status (graduate or undergraduate), major, GPA range, self-reported abilities, race, and gender. The coefficient of interest is β, which captures taste-based 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 employer fixed effects (μ_i). Standard errors in equation (1) are clustered by employer and worker. We also consider the role of worker beauty for employers' performance expectations; for this analysis, we simply replace W_{ij} with $E_i[P_j]$ in equation 1 and omit the performance rank from the right-hand side.

In order to determine whether employer beliefs about worker performance in a given task are correct on average, we test whether actual performance and attractiveness 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, T indexes the task, and t indexes the date of the session. We suppress the round subscripts for tractability. The variable P_{jT} is the performance in points of worker j in task T in round 1 or in round 2. Worker characteristics are captured by X_j'. The coefficient of interest is ρ, which captures the correlation between performance and attractiveness. In all specifications, we include a set of date fixed effects (σ_t). In both equations, we also allow the coefficient on the attractiveness rating to vary by attractiveness quintile to test for non-linear effects.

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, in 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 completes a project). If attractiveness is used as a signal of future performance, it should be less informative in the second round, after a worker's actual ability is observed more precisely. To test for this type of learning, we estimate equation (1) separately in round 1 and round 2. As an explicit test of whether employers update their beliefs and bidding behavior based on performance information, we then also include the worker's past performance in the round 2 regressions.

4. Results

We now explore the sources of the task-specific beauty premium found in Table 2 and examine whether it persists after employers learn about past performance, using the empirical framework outlined above. Any results not shown in the paper are available in the online appendix.

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 analysis and for the middle attractiveness quintiles in data entry. There is a beauty penalty for the most attractive workers in data entry.

Support for Result 1 comes from Table 3, which shows the relationship between the natural logarithm of the wage bid in round 1 and worker attractiveness, conditional on extensive controls. Columns 1 and 2 pool the data across tasks and include task fixed effects. Column 1 assumes that the relationship between the wage bid and the attractiveness rating is linear, while Column 2 breaks the attractiveness rating into quintiles to allow for nonlinearities. On average, a one-standard-deviation increase in attractiveness leads to a 16 percent increase in the wage offer. Workers whose beauty rating falls into the 4^{th} quintile (above-average looks) and the top attractiveness quintile receive significantly higher wage bids than workers in the bottom quintile.[7]

[TABLE 3 ABOUT HERE]

When we further decompose our analysis, we observe that the beauty premium varies by task. In particular, we do not find a significant effect of beauty on wage bids in either the data entry or the data analysis task, on average (Columns 3 and 5). In data analysis, moderately attractive workers (4^{th} quintile) receive a marginally higher wage than the workers in the bottom quintile (Column 4). However, the coefficient is not statistically significantly different from the coefficients on the 2^{nd}, 3^{rd}, or 5^{th} quintiles. In data entry, workers in the 2^{nd}, 3^{rd}, and 4^{th} attractiveness quintile receive significantly higher wage bids relative to the bottom quintile (Column 6), with the moderately attractive workers (4^{th} quintile) receiving the largest premium. Furthermore, an F-test shows that workers in the top attractiveness quintile receive significantly lower wage bids than workers in the 4^{th} quintile ($p = 0.006$), which suggests a beauty penalty for the most attractive workers. 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 27 percent. Column 8 shows that the beauty premium in the bargaining task is strongest for the top quintile (the most attractive workers).

The fact that the beauty premium is strongest 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

[7] An F-test reveals that the coefficients on the 4^{th} and 5^{th} quintiles are not statistically different from one another. These two coefficients are jointly significantly different from the coefficients on the 2^{nd} and 3^{rd} quintiles ($p = 0.019$ and $p = 0.008$, respectively).

premium. Because employer expectations about the relationship between attractiveness and performance should play a role in the relationship between attractiveness and wage offers, we next examine whether employers believe that more attractive workers are more productive in the three tasks.

Result 2: Employers incorrectly expect more attractive workers to be more productive in the bargaining task and correctly believe that attractiveness is unrelated to productivity in the other two tasks.

[TABLE 4 ABOUT HERE]

Table 4 shows the estimated relationship between an employer's performance expectation and worker attractiveness, conditional on the same controls as above. When we pool the data across tasks, we do not observe a significant linear relationship between worker attractiveness and employer performance prediction (Column 1). The most attractive workers (5th quintile) are expected to have a marginally significant performance advantage over those in the bottom quintile (Column 2). As we anticipated, employers do not expect relatively attractive workers to have a performance advantage in the data entry or the data analysis task (Columns 3-6). In fact, employers expect the most attractive workers to be less productive in data analysis relative to the other three attractiveness quintiles combined (Column 4, $p = 0.075$). This finding is consistent with the common stereotype that people who excel at analytical tasks are less attractive. For example, elementary school children presented with photos of 10 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 accordance with this stereotype, Deryugina and Shurchkov (2015) find that relatively 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.

On average, we do not find a significant relationship between beauty and expected performance in bargaining (Column 7), though it does emerge for workers rated above average in terms of attractiveness (Column 8). In particular, workers in the 5th quintile are expected to perform 18 percent better and workers in the 4th quintile are expected to perform 20 percent better than the workers in the bottom quintile. The coefficients on the above-average attractiveness quintiles are also jointly significantly higher than the coefficients on 2nd and 3rd quintiles ($p = 0.058$). This finding is consistent with the "beauty-related" nature of the task, because workers can see their opponent's photos during bargaining.

Table 5 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 worker's beauty rating or on the indicator that her beauty rating is in a given quintile. 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. The results are similar if we do not include these additional controls. The lack of a beauty performance advantage in our bargaining task differs from the previous literature that finds that attractive subjects outperform their less attractive counterparts in the ultimatum game – a simple bargaining game (Solnick and Schweitzer, 1999). Similarly, there are no beauty-based performance differentials in any of the other tasks or overall (Columns 1-6).

[TABLE 5 ABOUT HERE]

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 positive relationship between beauty and wage bids disappears in bargaining 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 6, 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, on average, and the point estimates are much smaller (Columns 1, 3, 5, and 7). When we allow the effects of attractiveness to be nonlinear, we still find a beauty premium for the workers of above-average attractiveness in the pooled data (Column 2). This is driven by the data entry task (Column 6), where workers in the 2nd and 4th quintiles of attractiveness receive higher wage bids than workers in the bottom quintile. Finally, beauty is no longer a significant determinant of the wage bid in either data analysis or bargaining (Columns 4 and 8).

[TABLE 6 ABOUT HERE]

As explained in Section 3, any residual relationship between beauty and wage bids can be interpreted as taste-based discrimination. Although we find such a relationship for two of the beauty quantiles in Column 6, this correlation is likely spurious, for two reasons. First, there is no clear theoretical reason why only workers in the 2nd and 4th quintiles should be preferred by employers, all else equal. Second, we would not expect taste-based discrimination to vary across tasks or rounds; however, we do not find a similar pattern in any of the other tasks, or, as we show later, in the second round.

The absence of taste-based discrimination 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 3–6 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.

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

Recall that Columns 5-8 of Table 2 document a marginally 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. Specifically, we estimate the effect of attractiveness on wage offers in the second round, with and without controlling for workers' first-round performance.

We hypothesize that, because we do not 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: In the second round, the beauty premium is absent from all tasks, on average. A beauty premium arises for attractive workers (4th and 5th quintiles) in data analysis.

Support for Result 4 comes from Table 7, which is the round two equivalent of Table 3. Like Columns 5-8 of Table 2, Columns 1, 3, 5, and 7 of Table 7 show the relationship between the natural logarithm of the wage bid in round 2 and worker attractiveness. However, we now condition on worker résumé characteristics. Controlling for worker characteristics eliminates the linear beauty premium in the second round on average: beauty is no longer correlated with wage bids in any of the tasks.

The beauty premium persists for workers in the 2nd and 4th quintiles of attractiveness when we pool the data (Column 2) and for workers in the 4th and 5th quintiles of attractiveness in data analysis (Column 4). We do not find a nonlinear relationship between beauty and wages in the other two tasks (Columns 6 and 8). In the online appendix, we show that controlling for the employer's prediction rank in the second round completely eliminates the remaining beauty premium in data analysis and overall, suggesting that the observed premiums are again due to statistical discrimination.

[TABLE 7 ABOUT HERE]

Result 5: In the second round, the beauty premium vanishes completely when the task is the same in both rounds. Past performance is an important determinant of wage bids, especially when the task in the second round is the same as in the first round.

Support for Result 5 comes from Table 8 which separates the effect of beauty on wage bids by whether the task was the same or different in the second round. When we do not control for performance in the first round (Columns 1, 3, 5, and 7), we find no evidence of a beauty premium. In fact, we find a beauty penalty in bargaining for attractive workers who faced a different task in the second round relative to comparably attractive workers whose first-round task was also bargaining (Column 7, $p = 0.015$ for difference between the two coefficients). Controlling for employer prediction rank in addition to past performance interacted with the first-round task eliminates this beauty penalty (estimates available in the online appendix). Although we are not able to explain it definitively, we hypothesize that the beauty penalty arises because more attractive workers perform slightly (though not significantly) worse in first-round data entry and data analysis. Employers may infer that more attractive workers will also perform worse when the second-round task is bargaining. Indeed, the point estimate for the effect of beauty on second-round prediction when bargaining in the second round is a different task is negative, though not statistically significant.

[TABLE 8 ABOUT HERE]

Columns 2, 4, 6, and 8 of Table 8 include controls for past performance interacted with the indicators for first-round task.[8] Overall, these specifications show a substantial amount of learning in the second round, as past performance is a significant predictor of wage bids in all specifications. Furthermore, Columns 4, 6, and 8 reveal that the significance of past performance in the employer's wage offer decision is greatest when the task in the second round is the same as the first-round task (coefficients bolded).

Once we properly control for past performance, we find that the beauty premium reemerges when we pool the data across tasks (Column 2), but only when the task in the second round is different from the first-round task. This result is consistent with our hypothesis about the informational value of beauty. 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 when the task is the same 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. However, when the task in the second round is different, information about past performance does not help resolve the uncertainty about worker ability in the new task, and appearance may again be utilized as a proxy for ability. Indeed, we find that past performance significantly affects employer beliefs about subsequent performance in the second round. Consequently, as shown in the online appendix, employer beliefs are no longer incorrect in bargaining or in any of the tasks.

4.3. Robustness Tests and Extensions

We check whether the results are robust to a variety of alternative specifications and explore several potential sources of heterogeneity in the beauty premium. First, linear specifications with employer bids, predictions, and worker performance in levels produce qualitatively similar results in both rounds.

Second, the beauty premium does not vary systematically by gender, although we do find a few significant differences. In the first round, employers expect more attractive males to perform better than the more attractive females in bargaining, but not in other tasks. This, however, does not translate into a differential beauty premium. In the second round, more attractive females are expected to perform better than the more attractive males in data entry, but not in the other tasks. As a result, we observe significantly higher wages offered to more attractive female workers relative to their male counterparts in data entry. This effect disappears once we control for the employer performance prediction which suggests that the observed premium was due to statistical discrimination. We do not observe any gender-specific effects of beauty on performance in any of the tasks or rounds.

Third, we explore the strategic implications of the binding of the maximum bid rule and find that the beauty premium in bargaining in the first round exists only when the total bid maximum does not bind.

Fourth, we consider another possible explanation for the beauty premium in the first round: more attractive individuals are more confident in the bargaining task (MR 2006). We find no positive relationship between beauty and the worker's prediction error – our measure of confidence – defined as the worker's own performance prediction minus that worker's actual

[8] Overall, we do not observe a systematic effect of task order on behavior in the second round, with two exceptions. First, when the round 1 task is data entry, employers predict slightly higher performance in second-round bargaining or data analysis. Second, workers perform significantly better in second-round data analysis when the first-round task is also data analysis, either due to learning or because the version of data analysis used following first-round data analysis was easier.

performance in a given task. Fifth, we do not find any systematic variation in the beauty premium or in the employer beliefs based on the order in which employers saw worker information: photo first or resume first. Detailed discussions of the five aforementioned analyses can be found in the online appendix.

Finally, our results also do not change substantively if we omit the middle or the top attractiveness quintile rather than the bottom quintile or if we exclude the two sessions that contained the two subjects who withdrew from our study.

5. Conclusion

Our results indicate that the beauty premium is highly context-dependent: while we find strong evidence for a beauty premium in a bargaining task, there is no beauty premium in data entry or data analysis. In fact, there is a beauty penalty for the most attractive workers in data entry. Both of these types of discrimination are entirely statistical and can be explained by biased beliefs about the performance of relatively attractive workers. We also find a strong learning effect: if worker performance is revealed and the task is repeated, discrimination disappears. This suggests that, in our setting, employers use attractiveness primarily as an imperfect signal of ability.

The laboratory experiment allowed us to control for worker characteristics (other than beauty) that are observable to the employer but that are often difficult, if not impossible, to include in observational studies. Our results show that controlling for these resume characteristics is important. For example, without them, we observe a persistent a beauty premium in bargaining. However, the second-round beauty premium in bargaining disappears just by including comprehensive worker controls.

Currently, employers are not prohibited from discriminating based on attractiveness either during the hiring process or in subsequent promotion decisions. Whether this is welfare-enhancing depends on numerous factors outside the scope of this paper, including society's equity concerns. However, one element of this question to which our paper *can* speak is the efficiency loss stemming from biased beliefs. If employers are consistently mistaken about the relative performance of more attractive people, then there will be an efficiency loss in the labor market. Because we do not find persistent biased beliefs in favor of more attractive people, our results suggest that this efficiency loss is likely to be small.

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 tastes if they expect to interact with them in person. 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. 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 such as ours, where subjects interact with each other largely through computers, are increasingly relevant outside of the laboratory.

Our results have two key implications about the extent to which learning mitigates the importance of beauty in the labor market. On one hand, the beauty premium may be smaller outside the laboratory, where employers are usually experienced in hiring and where feedback on worker productivity is available. On the other hand, our employers observe the subsequent performance of *all* workers, even those they did not hire. This should result in more learning in our experiment relative to the typical situation where the performance of workers who were not hired is not perfectly observed. The one exception is online labor markets, where proxies for performance (e.g., worker ratings) can be observed for all workers. Furthermore, performance information in our experiment is relatively precise and salient. Outside the laboratory, employers may learn about workers' true ability more slowly over time. Similar to Pallais (2014), who finds that firms hire too few inexperienced workers in general, our findings suggest that providing workers with more opportunities to formally demonstrate their ability would enhance efficiency in hiring, particularly in settings where statistically incorrect beliefs may be present.

References

Amiti, M., Davis, D. R., 2012. Trade, firms, and wages: theory and evidence. *The Review of Economic Studies*, 79(1), 1–36.

Andreoni, J. Petrie, R., 2008. Beauty, gender and stereotypes: evidence from laboratory experiments. *Journal of Economic Psychology*, 29(1), 73–93.

Arunachalam, R., Shah, M., 2012. The prostitute's allure: the return to beauty in commercial sex work. *The BE Journal of Economic Analysis and Policy*, 12(1).

Autor, D.H., Dorn, D., Hanson, G. H., 2013. The China syndrome: local labor market effects of import competition in the United States. *The American Economic Review*, 103(6), 2121-2168.

Ayres, I., 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.

Belot, M., Bhaskar, V., Van De Ven, J., 2012. Beauty and the sources of discrimination. *Journal of Human Resources*, 47(3), 851–872.

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

Berri, D. J., Simmons, R., Van Gilder, J., 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., Hamermesh, D., 1998. Beauty, productivity, and discrimination: lawyers' looks and lucre. *Journal of Labor Economics*, 16(1), 172–201.

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

Borland, J., Leigh, A., 2014. Unpacking the beauty premium: what channels does it operate through, and has it changed over time? *Economic Record*, 90(288), 17–32.

Bottomley, L., Parry, E., Brigade, S., Coley, L., Deam, L., Goodson, E., Kidwell, J., Linck, J, 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.

Card, D., Devicienti, F., Maida, A., 2014. Rent-sharing, holdup, and wages: evidence from matched panel data. *The Review of Economic Studies*, 81(1), 84–111.

Castillo, M., Petrie, R., 2010. Discrimination in the lab: Does information trump appearance? *Games and Economic Behavior*, 68(1), 50–59.

Castillo, M., Petrie, R., Torero, M., 2012a. Beautiful or White? Discrimination in group formation. Working Paper.

Castillo, M., Petrie, R., Torero, M., Vesterlund, L., 2012b. 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., 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.

Dermer, M., Thiel, D. L., 1975. When beauty may fail. *Journal of Personality and Social Psychology*, 31(6), 1168-1176.

Deryugina, T., Shurchkov, O., 2015. Does beauty matter in undergraduate education? *Economic Inquiry*, 53(2), 940-961.

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

Fershtman, C., Gneezy, U., 2001. Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, 116(1), 351–377.

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

Fletcher, J., 2009. Beauty vs. brains: Early labor market outcomes of high school graduates. *Economic Letters*, 105(3), 321–325.

Gächter, S., 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. S., 2006. Changing looks and changing "discrimination": The beauty of economists. *Economics Letters*, 93(3), 405–412.

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

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

Hamermesh, D. S., Parker, A., 2005. Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4), 369–376.

Harper, B., 2000. Beauty, stature and the labour market: A british cohort study. *Oxford Bulletin of Economics and Statistics*, 62, 771–800.

Kahneman, D., Slovic, P., Tversky, A. (Eds.) 1982. *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, United Kingdom.

Leigh, A., Susilo, T., 2009. Is voting skin-deep? Estimating the effect of candidate ballot photographs on election outcomes. *Journal of Economic Psychology*, 30(1), 61-70.

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

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

Pallais, A., 2014. Inefficient hiring in entry-level labor markets. *American Economic Review*, 104(11), 3565–3599.

Ponzo, M., Scoppa, V., 2012. The good, the bad, and the ugly: Teaching evaluations, beauty and abilities. Working paper.

Pope, D. G., Sydnor, J.R., 2011. What's in a picture? Evidence of discrimination from Prosper. com. *Journal of Human Resources*, 46(1): 53–92.

Rabin, M., 1998. Psychology and economics. *Journal of Economic Literature*, 36(1), 11–46.

Ravina, E., 2012. Love & loans: the effect of beauty and personal characteristics in credit markets. Working Paper, Available at SSRN: http://ssrn.com/abstract=1101647.

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

Ruffle B. J., Shtudiner, Z., 2014. Are good-looking people more employable? *Management Science*, in press.

Scholz, J.K., Sicinski, K., 2015. Facial attractiveness and lifetime earnings: Evidence from a cohort study. *Review of Economics and Statistics*, 97(1), 14–28.

Sen, A., Voia, M. C., Woolley, F. R., 2010. Hot or not: How appearance affects earnings and productivity in academia. Working paper.

Solnick, S. J., 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.

von Bose, C., 2013. Child stars vs. ugly ducklings: does adolescent attractiveness contribute to the beauty premium? The University of Texas at Austin mimeo.

Walker, W. R., 2013. The transitional costs of sectoral reallocation: Evidence from the Clean Air Act and the workforce. *The Quarterly Journal of Economics*, 128(4), 1787–1835.

Wilson, R. K., 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 Analysis		Data Entry		Bargaining	
	# Sessions	#Subjects	# Sessions	#Subjects	# Sessions	#Subjects
Round 1	15	120	15	120	15	120
Round 2	15	120	15	120	15	120
Round 1 = 2	8	64	7	56	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

The beauty premium by task in both rounds

Outcome variable:	Natural logarithm of employer wage bid							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks	Data An.	Data Entry	Barg.	All Tasks	Data An.	Data Entry	Barg.
	[26.5]	[23.2]	[27.7]	[28.5]	[27.4]	[22.7]	[30.1]	[28.5]
	ROUND 1				ROUND 2			
Attractiveness of worker	0.172**	0.128	0.104	0.241**	0.147*	0.069	0.172	0.210*
	(0.074)	(0.158)	(0.106)	(0.107)	(0.078)	(0.130)	(0.161)	(0.114)
Observations	712	236	236	240	712	240	236	236
R-squared	0.283	0.399	0.229	0.206	0.28	0.279	0.165	0.375

Notes: 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. All regressions include employer 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 and worker in parentheses. Significance levels: ** 5 percent, *** 1 percent.

Table 3

The beauty premium by task in round 1, conditional on worker characteristics

Outcome variable:	Natural logarithm of employer wage bid in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Analysis		Data Entry		Bargaining	
	[26.5]		[23.2]		[27.7]		[28.5]	
Attractiveness of worker	0.162**		0.136		0.007		0.265**	
	(0.065)		(0.132)		(0.105)		(0.111)	
Attractiveness quintiles:								
2nd		0.28		0.256		0.922**		0.996**
		(0.231)		(0.405)		(0.411)		(0.494)
3rd		0.139		0.354		0.709*		0.234
		(0.213)		(0.405)		(0.411)		(0.448)
4th		0.511**		0.628*		1.106**		0.759**
		(0.202)		(0.360)		(0.418)		(0.330)
Top attractiveness: 5th		0.445*		0.348		0.214		1.078***
		(0.227)		(0.376)		(0.384)		(0.401)
F-test p-value		[0.064]		[0.418]		[0.015]		[0.087]
Observations	712	712	236	236	236	236	240	240
R-squared	0.40	0.40	0.59	0.60	0.39	0.42	0.34	0.36

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 employer 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 and worker in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 4

Relationship between employer performance expectations and worker attractiveness by task 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 Analysis		Data Entry		Bargaining	
	[63.0]		[53.2]		[72.5]		[63.3]	
Attractiveness of worker	0.028		-0.004		-0.123		0.039	
	(0.017)		(0.019)		(0.172)		(0.026)	
Attractiveness quintiles:								
2nd		0.004		0.034		0.027		0.069
		(0.080)		(0.046)		(0.064)		(0.227)
3rd		-0.049		0.04		-0.111		-0.151
		(0.054)		(0.040)		(0.139)		(0.151)
4th		0.039		0.048		-0.04		0.197**
		(0.048)		(0.042)		(0.074)		(0.096)
Top attractiveness: 5th		0.076*		-0.016		-0.043		0.180**
		(0.042)		(0.055)		(0.042)		(0.075)
F-test p-value		[0.421]		[0.327]		[0.874]		[0.036]
Observations	712	712	236	236	236	236	240	240
R-squared	0.85	0.85	0.98	0.98	0.79	0.79	0.71	0.72

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 employer 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 and worker in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 5

Relationship between worker performance and worker attractiveness by task in round 1

Outcome variable:	Natural logarithm of worker performance in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Analysis		Data Entry		Bargaining	
	[64.2]		[35.8]		[89.5]		[64.2]	
Attractiveness of worker	0.002		-0.024		-0.117		-0.054	
	(0.048)		(0.053)		(0.085)		(0.168)	
Attractiveness quintiles:								
2nd		0.125		0.101		-0.132		0.003
		(0.173)		(0.175)		(0.334)		(0.729)
3rd		-0.017		0.033		-0.342		-0.106
		(0.210)		(0.207)		(0.301)		(0.627)
4th		0.212		-0.038		-0.312		0.781
		(0.180)		(0.169)		(0.381)		(0.765)
Top attractiveness: 5th		0.013		-0.032		-0.491		0.025
		(0.184)		(0.210)		(0.338)		(0.589)
F-test p-value		[0.590]		[0.817]		[0.195]		[0.624]
Observations	178	178	59	59	59	59	60	60
R-squared	0.39	0.40	0.68	0.69	0.54	0.59	0.50	0.54

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 in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 6

The beauty premium by task in round 1, conditional on worker characteristics and employer prediction rank

Outcome variable:	Natural logarithm of employer wage bid in round 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Analysis		Data Entry		Bargaining	
Attractiveness of worker	0.093		0.084		0.002		0.097	
	(0.060)		(0.129)		(0.095)		(0.126)	
Attractiveness quintiles:								
2nd		0.301		0.155		0.858*		0.463
		(0.215)		(0.333)		(0.439)		(0.445)
3rd		0.131		0.373		0.517		0.339
		(0.180)		(0.344)		(0.424)		(0.348)
4th		0.431**		0.471		0.912**		0.263
		(0.169)		(0.324)		(0.389)		(0.376)
Top attractiveness: 5th		0.315		0.206		0.35		0.33
		(0.202)		(0.345)		(0.407)		(0.409)
F-test p-value		[0.108]		[0.532]		[0.015]		[0.715]
Observations	712	712	236	236	236	236	240	240
R-squared	0.54	0.54	0.66	0.66	0.53	0.55	0.54	0.54

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 on the coefficients of attractiveness quintiles are reported in brackets below the estimates. All regressions include employer's performance prediction rank, employer 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 and worker in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Table 7

The beauty premium by task in round 2, conditional on worker characteristics

Outcome variable:	Natural logarithm of employer wage bid in round 2							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Analysis		Data Entry		Bargaining	
	[27.4]		[22.7]		[30.1]		[28.5]	
Attractiveness of worker	0.109		0.165		0.212		0.023	
	(0.080)		(0.163)		(0.203)		(0.119)	
Attractiveness quintiles:								
2nd		0.578*		0.914		0.699		0.632
		(0.300)		(0.666)		(0.592)		(0.610)
3rd		0.071		0.598		0.22		0.218
		(0.277)		(0.891)		(0.506)		(0.398)
4th		0.594**		1.194**		0.838		0.521
		(0.271)		(0.587)		(0.669)		(0.493)
Top attractiveness: 5th		0.393		0.986*		0.597		0.212
		(0.263)		(0.573)		(0.670)		(0.385)
F-test p-value		[0.102]		[0.175]		[0.599]		[0.808]
Observations	712	712	240	240	236	236	236	236
R-squared	0.34	0.35	0.48	0.50	0.33	0.34	0.49	0.50

Notes: Round 2 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 employer 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 8

The beauty premium by task type in round 2, conditional on worker characteristics and past performance

Outcome variable:	Natural logarithm of employer wage bid in round 2							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Tasks		Data Analysis		Data Entry		Bargaining	
Attractiveness of worker & same task in round 2	0.154	0.081	0.098	-0.063	0.046	-0.056	0.130	0.159
	(0.118)	(0.100)	(0.207)	(0.164)	(0.417)	(0.283)	(0.123)	(0.101)
Attractiveness of worker & different task in round 2	0.065	0.156*	0.208	0.111	0.297	0.352	-0.480*	-0.230
	(0.106)	(0.082)	(0.209)	(0.112)	(0.241)	(0.238)	(0.254)	(0.209)
F-test p-value (equality)	[0.572]	[0.547]	[0.860]	[0.385]	[0.536]	[0.245]	[0.015]	[0.074]
Log performance in round 1 & Data Analysis in round 1		1.482***		**2.741****		1.075		1.735*
		(0.446)		**(0.588)**		(1.193)		(0.948)
Log performance in round 1 & Data Entry in round 1		1.633***		1.440***		**2.250****		1.085*
		(0.315)		(0.394)		**(0.527)**		(0.645)
Log performance in round 1 & Bargaining in round 1		0.878***		0.953***		1.376***		**0.654****
		(0.130)		(0.252)		(0.460)		**(0.121)**
Observations	712	712	240	240	236	236	236	236
R-squared	0.34	0.48	0.48	0.61	0.34	0.42	0.51	0.60

Notes: Round 2 data only. The attractiveness coefficients 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 for equality of the coefficients on the attractiveness interactions with task type are reported in brackets below the estimates. All regressions include employer fixed effects, indicators for whether the second-round task was the same as the first-round task, 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 and worker in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.