WORKING PAPER · NO. 2019-86

Inaccurate Statistical Discrimination

J. Aislinn Bohren, Kareem Haggag, Alex Imas, and Devin G. Pope

JUNE 2019

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

INACCURATE STATISTICAL DISCRIMINATION

J. Aislinn Bohren
Kareem Haggag
Alex Imas
Devin G. Pope

June 2019

We thank Emir Kamenica, Alex Frankel, and seminar participants at Harvard Business School, Stanford University, UCLA, University of Chicago, University of Melbourne, University of Pennsylvania, University of Sydney, and University of Virginia for helpful comments and suggestions. Jihong Song provided excellent research assistance.

Inaccurate Statistical Discrimination
J. Aislinn Bohren, Kareem Haggag, Alex Imas, and Devin G. Pope
June 2019
JEL No. D90,J71

ABSTRACT

Discrimination has been widely studied in economics and other disciplines. In addition to identifying evidence of discrimination, economists often categorize the source of discrimination as either taste-based or statistical. Categorizing discrimination in this way can be valuable for policy design and welfare analysis. We argue that a further categorization is important and needed. Specifically, in many situations economic agents may have inaccurate beliefs about the expected productivity or performance of a social group. This motivates our proposed distinction between accurate (based on correct beliefs) and inaccurate (based on incorrect beliefs) statistical discrimination. We do a thorough review of the discrimination literature and argue that this distinction is rarely discussed. Using an online experiment, we illustrate how to identify accurate versus inaccurate statistical discrimination. We show that ignoring this distinction – as is often the case in the discrimination literature – can lead to erroneous interpretations of the motives and implications of discriminatory behavior. In particular, when not explicitly accounted for, inaccurate statistical discrimination can be mistaken for taste-based discrimination, accurate statistical discrimination, or a combination of the two.

J. Aislinn Bohren
Carnegie Mellon University and
University of Pennsylvania
4765 Forbes Avenue
Pittsburgh, PA 15213
abohren@gmail.com

Kareem Haggag
Social and Decision Sciences
Carnegie Mellon University
Porter Hall 208-H
5000 Forbes Avenue
Pittsburgh, PA 15213
kareem.haggag@cmu.edu

Alex Imas
Social and Decision Sciences
Carnegie Mellon University
Porter Hall 319C
500 Forbes Ave.
Pittsburgh, PA 15213
aimas@andrew.cmu.edu

Devin G. Pope
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
devin.pope@chicagobooth.edu

1 Introduction

Economists define discrimination as differential treatment of otherwise identical individuals from different social groups (i.e. race, gender, age, etc.). Discrimination has been shown to be prevalent in labor markets, housing markets, credit markets, and online consumer markets among others. In addition to observational data analysis, a variety of empirical techniques have been used to document discrimination, including audit studies, own-group bias, and correspondence studies (for recent reviews of the discrimination literature, see Bertrand and Duflo 2017 and Charles and Guryan 2011).

In addition to establishing the existence of discrimination, economists typically categorize discrimination as one of two types. The first type, taste-based discrimination (Becker 1957), posits that an individual or firm has animus towards members of a particular group, and therefore may choose to discriminate against them because he receives disutility from providing services to or interacting with members of the group. The second type, statistical discrimination, suggests that discrimination may occur against members of a particular group because productivity is unobserved and the group is perceived to have a lower average productivity (Phelps 1972; Arrow 1973) or a different variance (Aigner and Cain 1977).

Distinguishing between these two types of discrimination is important for several reasons. First, designing an effective policy intervention to reduce discrimination crucially depends on the source of discrimination. Second, welfare/efficiency analyses differ for each type of discrimination. For example, statistical discrimination is sometimes referred to as "efficient discrimination," in that it is the optimal response to a signal-extraction problem. Importantly, this premise of efficiency relies on the assumption that statistical discrimination stems from rational expectations, or correct beliefs, about the group distributions of the relevant outcome. Finally, the extent to which competitive markets will eliminate discrimination depends on its source.

In this paper, we argue that in many situations, an individual's beliefs about the productivity of different social groups may be inaccurate. We refer to discrimination that stems from inaccurate beliefs as *inaccurate statistical discrimination*. Just as it is important to distinguish between taste-based and statistical discrimination for policy design and welfare analysis, we show that it is equally important to separate inaccurate from accurate statistical discrimination. For example, if discrimination stems from inaccurate beliefs, an effective policy response could be to provide individuals with information about the correct distributions (as Jensen 2010 did in the case of inaccurate beliefs about the returns to education).

We discuss two broad sources for inaccurate beliefs. First, psychological biases and heuristics may lead to inaccurate beliefs. Psychologists have long explored and tried to understand why inaccurate stereotypes exist (Schneider et al. 1979; Judd and Park 1993, Hilton and Hippel 1996). Bordalo et al. (2016) provide a model for inaccurate stereotype formation based on the representativeness heuristic. Biased beliefs can also form and persist in a dynamic social learning setting when individuals have incorrect models of how others evaluate workers (e.g. Bohren et al. 2019). Second, inaccurate beliefs may simply be due to a lack of information. A completely rational actor may lack the relevant information necessary to form correct beliefs. For example, an employer may have unbiased prior belief about the average productivity of individuals from two different social groups, but be unaware that there is positive selection into the job application process for members of one group. Thus, an employer may not hire members of a particular group because he/she has an inaccurate belief due to limited information about the selection process. Over time, learning will mitigate inaccurate beliefs in some settings. But in other situations, there will be little or no feedback on the decisions being made, leading to learning traps in which inaccurate beliefs persist. Whether due to psychological bias or lack of information, inaccurate beliefs have been shown to exist in a variety of important domains, including the value of human capital formation (Jensen 2010), the prevalence of affirmative action (Kravitz and Platania 1993), and the extent of wealth inequality in the US (Norton and Ariely 2011).

By not allowing for the possibility of inaccurate beliefs, we demonstrate that the discrimination literature may mistakenly interpret evidence of inaccurate statistical discrimination as taste-based and/or accurate statistical discrimination. For example, a paper that measures productivity outcomes may find that discrimination against a particular group was not statistically warranted, as different groups had identical distributions of productivity, and therefore argue that it must be taste-based. However, an alternative explanation is that decision-makers had inaccurate beliefs and were inaccurately statistically discriminating, rather than discriminating due to animus.

To understand the prevalence of these mistaken interpretations, we review the empirical literature on discrimination. We find 105 papers published in the top economics journals between 1990 and 2018 that test for evidence of discrimination. Of these papers, 61.9% of them discuss statistical vs. taste-based discrimination and 46.7% of them test between the sources. Only 10.5% of them discuss the possibility of inaccurate beliefs, and 4.8% test for inaccurate beliefs. In nearly all of these cases, we argue that inaccurate statistical discrimination is a reasonable alternative to the interpretation chosen by the authors.[1]

[1]The line between inaccurate beliefs and taste-based discrimination may sometimes become a bit blurry.

To demonstrate how inaccurate beliefs influence patterns of discrimination, we run an online hiring experiment. We first have 589 participants (the "workers") take a 50-question math quiz. These workers differ in their country of origin (US vs. India), gender, and age. A different set of 577 participants (the "employers") are shown the profiles of 20 participants who took the math quiz and are asked the maximum they would be willing to pay in order to hire each participant (using a Becker-DeGroot-Marschak mechanism). If an employer hires a worker, the worker earns the wage and the employer is paid proportional to how many math questions the worker answered correctly.

Our first finding is that employers discriminate. Americans and females receive systematically lower wage offers than Indians and males (we find no statistically significant difference in wages offered to workers of different ages). Given the conventional categories of discrimination, this finding has two potential explanations. Employers may have animus towards Americans/females relative to Indians/males, and offer them lower wages because they do not want them to receive money. Alternatively, employers may statistically discriminate and offer Americans/females a lower wage because they believe that these groups have lower expected productivity than Indians/males.

One method often used in the discrimination literature to distinguish between statistical and taste-based discrimination is to measure the productivity of outcomes and compare them across groups. This determines whether the discrimination was "warranted" from a statistical perspective. If the difference in distributions by group can explain the observed level of discrimination, then statistical discrimination is at play, and otherwise, discrimination is taste-based. In our experiment, we find that Americans and Indians perform equally well on the math quiz, while females significantly underperform. Thus, using the standard approach to study discrimination, one would argue that the discrimination we observe against Americans stems from a taste-based source, as it was not warranted from a statistical perspective. Further, because we observe a level of discrimination against women that is too small relative to the large differences in productivity, this approach would also conclude that there is taste-based discrimination against men.

An alternative explanation for the discrimination that we find is that individuals have no animus towards a particular group, but rather, have inaccurate beliefs about how well these

What if individuals develop inaccurate beliefs because they have animus towards members of a particular group? Perhaps at times decision makers justify their animus with beliefs that deep down inside they know are incorrect. We agree that this is likely to be the case in many situations. However, we also think these things are separately identifiable. For example, if decision makers are provided with credible information, those that have inaccurate beliefs should change their behavior. Those whose inaccurate beliefs are merely masking an underlying animus are unlikely to change their behavior.

different groups performed on the math test.[2] To test for this alternative source, we elicit the beliefs of the employers. We find that employers mistakenly thought that Indians would perform much better on the math test than Americans and that females would only slightly underperform relative to males. Accounting for these inaccurate beliefs turns the prior conclusions on their head. What would have been diagnosed as taste-based discrimination *in favor of* Indians appears to be a combination of inaccurate statistical discrimination and taste-based discrimination *against* Indians. Similarly, a large portion of the gender gap is explained by inaccurate statistical discrimination.

As mentioned above, a potential policy intervention to mitigate inaccurate statistical discrimination is to provide individuals with information about how the distribution of an outcome varies by group. We provide the employers in our experiment with information about how the average performance on the math test varied by gender, nationality, and age. After receiving this information, participants were asked to make wage offers to 10 additional potential workers. We find that employers significantly changed their wage offers in the direction of correcting their inaccurate beliefs (less discrimination against Americans and more discrimination against females).

We view our experiment as a tool to illustrate the potential for erroneous conclusions when claiming to identify statistical versus taste-based discrimination without considering the possibility of inaccurate beliefs, rather than a domain for discrimination that is of inherent interest on its own. Our experiment also demonstrates the importance of understanding beliefs in the relevant market when choosing a policy intervention to mitigate discrimination, and the potential for policy interventions to effectively change beliefs.[3] Building on the nascent literature that discusses the possibility of discrimination due to inaccurate beliefs (see section 2 for a review), our results flesh out the implications for this additional source of discrimination, and provide an illustrative example of how allowing for inaccurate beliefs impacts the interpretation of experimental and observational findings.

The paper proceeds as follows. Section 2 presents a survey of the economics literature on discrimination. Section 3 presents the experimental design, analysis, and interpretation. Section 4 concludes.

[2]These inaccurate beliefs could be due to a psychological bias or simply a lack of information about how good/motivated people who are recruited to do this online task happen to be at math.

[3]Though see the end of Section 3 for some strong caveats on porting the information intervention in this experiment to the field.

2 Literature Review

We conducted a systematic survey of the economics literature on discrimination in order to determine: (1) how often papers seek to distinguish between taste-based and belief-based (statistical) sources of discrimination; (2) how often papers seek to distinguish between accurate and inaccurate beliefs for belief-based sources of discrimination. Most papers that met our inclusion criteria (outlined below) found evidence of discrimination: 102 out of 105 papers, or 97.1% documented evidence for discrimination against at least one group that was considered in the paper. The majority of papers (61.9%) discussed the source of discrimination as being driven by either preferences (taste-based) or beliefs (statistical), and nearly half of the papers (46.7%) attempted to distinguish between these two sources. However, very few papers discussed the possibility that beliefs may be inaccurate (10.5%), and fewer still examined whether beliefs were accurate or inaccurate (4.8%). Table 1 summarizes these findings.

Table 1: Summary of Literature Survey on Discrimination

	All: 1990 - 2018		**Recent: 2014 - 2018**	
	# Papers	**Percent**	**# Papers**	**Percent**
Total Papers	105	100.0%	31	100.0%
Evidence of Discrimination	102	97.1%	31	100.0%
Discuss taste-based versus statistical source	65	61.9%	23	74.2%
Test for taste-based versus statistical source	49	46.7%	16	51.6%
Discuss accurate versus inaccurate beliefs	11	10.5%	5	16.1%
Test for inaccurate beliefs	5	4.8%	2	6.5%
Measure beliefs	6	5.7%	3	9.7%

We classified papers as “discuss taste-based versus statistical source” if preference versus belief-based motives for the documented discrimination were discussed in the text, and as “test for taste-based versus statistical source” if the paper either explicitly tested between different models of preference versus belief-based discrimination or implicitly tested the predictions of a belief-based model while taking the taste-based model as the null hypothesis. If a paper mentioned inaccurate or biased beliefs as a potential source of discrimination, it was classified as “discuss accurate versus inaccurate beliefs.” Papers that tested whether inaccurate beliefs could be driving discrimination, either by directly eliciting beliefs or through other tests, were classified as “test for inaccurate beliefs.” Finally, papers that elicited beliefs were classified as “measure beliefs.” Three of the six papers in this category did not test whether these elicited beliefs were accurate.

Of the five papers that tested for inaccurate beliefs, three papers did so by directly eliciting beliefs (List 2004, Hedegaard and Tyran 2018,Mobius and Rosenblat 2006) and two did so by either comparing behavior across different contexts (Fershtman and Gneezy 2001) or using theoretical arguments (Arnold et al. 2018). List 2004 finds discrimination against minorities in a marketplace for sports cards. He elicits perceptions of reservation values from a separate group of card dealers and shows that these perceptions qualitatively match the distributional differences observed in the experiment. Experienced dealers are most accurate and exhibit the most discrimination. This suggests that the discrimination observed in that study is statistical. Hedegaard and Tyran (2018) examine the "price of prejudice" using a field experiment where workers can forgo potential earnings in order to work with an in-group member instead of an out-group member. They find evidence for animus workers are willing to forgo 8 percent of earnings to work with a member of the same ethnic group. They elicit productivity beliefs from a separate group of workers and show that these beliefs are qualitatively correct. This suggests that incorrect beliefs are not driving the observed discrimination. Mobius and Rosenblat (2006) was the only paper in our survey that documented inaccurate beliefs. They examine the beauty premium in an experimental principal-agent market and show that more attractive people are offered higher wages but are not more productive. By eliciting beliefs, the authors demonstrate that this discrimination is driven by an incorrect perception of productivity differences.

Method. In this section, we outline the method that we used to determine which papers to include in the survey and the data that we collected for each paper.

Inclusion Criteria. We focused on empirical papers published between 1990 and 2018 in the following journals: American Economic Journal: Applied, American Economic Journal: Policy, American Economic Review (excluding the Papers Proceedings issue), Econometrica, Journal of the European Economic Association, Journal of Labor Economics, Journal of Political Economy, the Quarterly Journal of Economics, Review of Economic Studies, and Review of Economics and Statistics. We acknowledge that the economics literature on discrimination includes important contributions from other journals. We restricted attention to these ten journals as a representative sample in order for the scope of the survey to include a manageable number of papers.

We proceeded in two steps to determine whether to include a paper published in the relevant time frame and journals. First, in each journal, we searched for all empirical papers that had at least one of the following search terms in the title:

{*discrimination, prejudice, bias, biases, biased, disparity, disparities, stereotype, stereotypes, premium*}
or at least one of the following search terms in the abstract:
{*discrimination, prejudice*}
or at least one of the following search terms from each list in the abstract:
{*racial, race, gender, sex, ethnic, religious, beauty*} AND {*bias, biased, disparity, stereotype, stereotypes, premium*}.
Second, we restricted attention to papers that attempted to causally document differential treatment of individuals based on their group identity. This eliminated papers on unrelated topics, including the industrial organization literature on *price* discrimination, the financial literature on the *risk* premium, theoretical models, and the experimental literature that documents behavioral differences such as gender differences in risk preferences.[4]

Data Collection. For each paper that met our inclusion criteria, we recorded the following information: data source (laboratory experiment, field experiment, audit study, observational data study, other), empirical method (reduced form analysis, structural analysis), group identity of interest (race, gender, ethnicity, religion, sexuality, class/income, other), domain of study (labor market, legal, education, financial, consumer purchases – non-financial, evaluations, other), measure of discrimination (i.e. difference in call back rates), whether the paper distinguishes between taste-based and statistical discrimination, whether the paper distinguishes between accurate and inaccurate statistical discrimination, whether discrimination was documented, whether the study identified the source of discrimination, and whether the study measured beliefs about an individuals predicted attribute by group identity.

Summary Statistics. We found 105 papers that met our inclusion criteria. Table 2 lists the number of papers broken down by journal and decade of publication.

[4] We also excluded some papers that met our objective criteria but which we viewed as not relevant to the spirit of the exercise. More specifically, we excluded papers that could not be classified as either a Yes or No for the criteria outlined in Table 1. For example, Gneezy et al. (2003) examine behavioral differences between men and women but do not study discrimination per se. Similarly, Cameron and Heckman (2001) examine the extent to which the racial and ethnic gap in college attendance can be explained by long-run versus short-run factors but do not address discrimination.

Table 2: Publications by Journal and Decade

	Number of Papers			
	1990-99	**2000-09**	**2010-2018**	**Total**
AEJ: Applied	0	1	7	8
AEJ: Policy	0	0	2	2
AER	4	7	6	17
EMA	0	0	0	0
JEEA	0	1	1	2
JLE	2	8	12	22
JPE	2	6	1	9
ReStud	1	2	3	6
ReStat	5	6	11	22
QJE	4	4	9	17
Total	**18**	**35**	**52**	**105**

Out of the papers surveyed, 11 conducted audit or correspondence studies, 7 conducted another type of field experiment, 3 conducted a laboratory experiment and 84 analyzed observational data.

Table 3: Publications by Method

Method	**# Papers**
Audit/Correspondence Study	11
Other Field Experiment	7
Laboratory Experiment	3
Observational Data	84

Discrimination was studied for a variety of group identities and in a variety of domains. The most frequent group identities were race (58 papers) and gender (37 papers), followed by physical traits / appearance (7 papers) and ethnicity (6 papers). The most frequent domain was labor markets (58 papers), followed by legal contexts (12 papers), education (9 papers), non-financial consumer markets (6 papers) and financial markets (5 papers). Table 4 summarizes the papers by group identity and domain. Some papers in the survey studied multiple group identities or domains; therefore, some papers are counted in multiple rows of the table.

Table 4: Type and Domain of Discrimination

	Total Papers	Evidence of Discrimination	
	#	#	Percent
Group Identity			
Race	58	56	96.6%
Gender	37	35	94.6%
Ethnicity	6	6	100.0%
Religion	1	1	100.0%
Sexuality	1	1	100.0%
Class/Income	1	1	100.0%
Physical Traits / Appearance	7	7	100.0%
Other	5	5	100.0%
Domain of Discrimination			
Labor Market	58	57	98.3%
Legal	12	12	100.0%
Education	9	9	100.0%
Financial	5	4	80.0%
Consumer Markets (not financial)	6	6	100.0%
Other	17	16	94.1%

3 Hiring Experiment

We have argued that it is necessary to elicit beliefs in order to identify the relative importance of (accurate) statistical versus taste-based motives for discrimination. We illustrate this point in a stylized hiring experiment that allows us to perform an accounting exercise typical of the outcomes-based tests common in the literature and to elicit relevant beliefs. We show that actual beliefs are inaccurate, violating the rational expectations assumption typically made in the statistical discrimination literature and altering the conclusions drawn from the exercise, compared to those that would be drawn under rational expectations. We also test a simple intervention to correct inaccurate beliefs: we provide individuals with information on the distributions of productivities by groups and conduct an additional hiring task. This measures the residual importance of taste-based discrimination and demonstrates the effectiveness of this type of informational intervention.

3.1 Experimental Design

Our experimental design includes two separate, pre-registered surveys: (1) a work task (math quiz) performed by 589 Amazon Mechanical Turk subjects (MTurkers), who comprise the prospective workers for the second survey, (2) a hiring task in which each of 577 different

MTurkers, who comprise the employers, stated a wage (willingness to pay)) for 20 prospective worker profiles.[5] The second survey also contains a belief elicitation and an information intervention followed by a second hiring task. We describe the experimental design and provide summary statistics below.

Survey 1 (Work Task): We recruited 589 subjects from MTurk on February 23, 2018 for the first survey.[6] The survey was posted with the title "Math Questions and Demographics" and the description "A 20-minute task of answering math questions." We paid \$2 (i.e. a projected \$6/hour wage) and recruited a subject pool of 392 from the United States and 197 from India, all of whom had completed at least 500 prior tasks and had an 80% or higher approval rate for these tasks.[7] After starting the survey, subjects were informed that they would first answer demographic questions and then answer 50 multiple choice math questions. They were told that their performance would not affect their payment, and were asked not to use a calculator or any outside help, but just to do their best. This was followed by seven questions that provided the information used for their profiles in the second survey: favorite color, favorite movie, coffee vs. tea preference, age, gender, favorite subject in high school, and favorite sport. The math test included a mix of arithmetic (e.g. "$5 * 6 * 7 =?$"), algebra (e.g. "If $(y + 9) * (y^2 - 121) = 0$, then which of the following cannot be y?"), and more conceptual questions (e.g. "Which of the following is not a prime number?").[8] Finally, subjects were thanked for their participation and informed that they may receive a small bonus based on a different experiment, for reasons unrelated to their performance on the task. We describe the basis for such bonuses in the description of Survey 2.

The purpose of the first survey was to create a bank of "workers" who could be hired by the "employers" in the second survey. This novel design has several advantages over the existing paradigms for studying discrimination in the field. First, in contrast to correspondence studies, we did not employ deception at any point all profiles shown to employers corre-

[5]We pre-registered the study on AsPredicted.org. There are two minor differences between the pre-registration plan and the actual study. First, we pre-registered that we would recruit 400 employers in the hiring task survey, but decided to recruit closer to 600. Second, we did not pre-register sample restrictions due to completing the task too quickly or slowly. We dropped 12 subjects in the work task survey and 5 in the hiring task survey due to these restrictions.

[6]We received 604 responses in total, but dropped 12 responses that corresponded to the top 1% (¡227 seconds) and bottom 1% (¿3274 seconds) in terms of survey duration. Of the remaining 592 responses, we dropped 3 whose Qualtrics survey responses could not be matched to their MTurk records, leaving 589 final respondents.

[7]This geographic restriction is based on the addresses MTurkers used to register on Amazon. The survey was posted as two tasks on MTurk, with one only eligible for Indian workers and one only eligible for U.S. workers.

[8]The full survey is available in the Online Appendix.

sponded to actual workers who would in fact be paid as described in the following paragraph. However, similar to a correspondence study, we were able to control the information seen by an employer about a prospective worker by constructing worker profiles that included information that is ostensibly relevant for animus and/or beliefs about productivity (e.g. age, gender, and nationality), as well as other irrelevant information (e.g. tea preference). The irrelevant information serves as a placebo test and ensures that the relevant information is not the only salient information provided to the employer (this mimics the irrelevant information contained on a CV). Finally, instead of the coarse measures of discrimination used in many other studies (e.g. callback or stop rates), we elicit relatively continuous and precise measures of productivity and discrimination that are tightly linked.

Survey 2: We recruited 577 different MTurk subjects on February 26, 2018. We used the same hiring criteria as the first survey (392 from U.S., 185 from India, >= 80% approval rate).[9] The survey was posted with the title "20-Minute Survey about Decision-Making" and the description "20-Minute Survey about Decision-Making." We paid \$2 (i.e. a projected \$6/hour wage). Subjects were first asked to report their gender, age, and education level. Subjects were then presented with the first hiring task portion of the survey.

First Hiring Task: We informed subjects that we had previously paid other subjects ("workers") to answer 50 math questions, showed them five examples of the math questions, and told them that on average, participants answered 36.95 out of 50 questions correctly. They were then told that they would act as an employer and hire one of these workers by stating a wage (paid as a bonus to the worker). In return, they would receive a payment based on how many questions their hired worker answered correctly. This was followed by a more detailed description of the assignment. Each "employer" would view 20 profiles of potential workers and state the highest wage (between 0 to 50 cents) they were willing to pay to each worker. The employer would be paid 1 cent for each question answered correctly by the hired worker. We next described the mechanism (Becker-Degroot-Marschak) used to assign payment. We would randomly select a profile from the 20 potential workers: We would then draw a random number from 0 to 50. If the wage the employer stated for the worker was equal or greater than that number, then the worker would receive the random number as a bonus and the employer would receive a "profit" equal to the workers performance minus the random number. If instead the employer stated a wage for the worker that was lower than the random number, then neither the worker nor the employer would receive a payment.

[9]We recruited 587 subjects in total, but dropped 7 whose surveys were completed in under 300 seconds and 3 whose stimuli (the profiles they evaluated) could not be matched to the first survey.

To ensure comprehension, we showed subjects an example profile (see Figure 1) and stated wage. We gave examples of actual performance and randomly generated numbers that would produce positive profit, negative profit, and no hiring. Having highlighted the possibility of negative profit, we then noted that all employers would automatically be paid a $0.50 bonus in addition to any money made through the hiring task, so that no employers would owe money. Finally, we ran a comprehension check with the same example profile, a specific wage (43), a random number (18), and an actual performance (10). We required the employer to correctly state how many cents they would have to pay the worker (18) and how many cents the employer would be paid before subtracting off the amount they would pay the worker (10).[10] Finally, employers were presented with a second wage (15), and answered the same questions. They were then presented with 20 profiles, each randomly selected with replacement from the bank of 589 profiles produced by the first survey.

Figure 1: Example Profile Used in First Hiring Task Description

Country:	United States
Gender:	Female
Age:	63
Favorite High School Subject:	English
Favorite Sport:	Gymnastics
Favorite Color:	Sea Green
Favorite Movie:	Overboard
Prefers Coffee/Tea:	Tea

Belief Elicitation Task: Next, subjects were randomly assigned to one of two different conditions: an incentivized or un-incentivized belief elicitation. Across both conditions, subjects were reminded that the full sample answered 36.95 out of 50 questions correctly. They were then asked to answer six questions of the form, “On average, how many math questions out of 50 do you think X answered correctly?” where X corresponded to the groups “women”, “men”, “people from the United States”, “people from India”, “people below or at the age of 33,” and “people above the age of 33.”

In the incentivized condition, prior to the six questions, subjects were told that they

[10] Entering an incorrect an answer would generate a pop-up with Wrong Answer and restrict the individual from moving to the next page.

could earn a significant bonus for an accurate prediction. One of the six questions would be randomly selected and they would be paid \$5 minus their deviation from the question (bounded below by \$0). For example, if they answered 40 and the true average was 37, they would receive a \$2 bonus. Finally, they were asked to "please answer the questions as carefully as possible so that you can potentially win a large bonus."

Information Intervention & Second Hiring Task: After completing the belief elicitation, subjects were shown the correct answer for all six groups: women (35.28), men (38.32), people from the U.S. (37.14), people from India (36.58), people below or at the age of 33 (37.10), and people above the age of 33 (36.79). Following this information, we stated, "Now that you have learned those facts, we would like you to work on 10 more profiles." We noted that, as in the first hiring task, we would randomly select one profile and a number, and pay bonus and wages accordingly (with an additional \$0.50 automatic bonus to ensure no negative payments). After employers reviewed the 10 additional worker profiles, we thanked them for their participation, noted that we would calculate bonuses and pay them within a week, and allowed subjects the option to leave comments.

Summary Statistics: Table 5 provides summary statistics for the full sample of subjects that completed surveys 1 and 2 (Column 1), as well as these statistics for each of the 6 demographic groups used in the second survey. On average, the work task (survey 1) took subjects 19 minutes to complete, while the hiring task took 23 minutes. There is variation in this timing across groups. Subjects from the U.S. took an average of 19 minutes to complete the hiring task, while subjects from India took 31.60 minutes; a difference also reflected in their median times (15.8 vs. 25.6). Another large difference between the U.S. and India samples is the average age of participants; the average Indian subject in the work task is approximately 8 years younger than the average American subject. This gap shrinks to 4 years for the hiring task. The Indian sample also skews more male than the U.S. sample (68.5% vs. 48.2% and 76.8% vs. 51.4% for survey 1 and 2, respectively) and is more likely to have a college education or above (90.3% vs. 56% in survey 2; the question was not asked in survey 1). While we primarily focus on simple comparisons between each demographic group, these observed differences motivate our use of multivariate regressions as well.

Table 5: Summary Statistics

	Total	Male	Female	US	India	Under 33	Over 33
Panel A: Worker							
Trivia Score	36.95	38.32	35.28	37.14	36.58	37.10	36.79
	(8.727)	(8.519)	(8.704)	(8.933)	(8.311)	(8.549)	(8.939)
Survey Duration (Minutes)	18.82	19.03	18.56	16.19	24.04	20.25	17.18
	(10.39)	(10.52)	(10.25)	(8.118)	(12.31)	(11.82)	(8.199)
Prefer Tea (Yes = 1)	0.394	0.380	0.411	0.370	0.442	0.424	0.360
	(0.489)	(0.486)	(0.493)	(0.483)	(0.498)	(0.495)	(0.481)
Age (Worker)	35.89	35.30	36.62	38.55	30.61	27.38	45.61
	(11.57)	(11.27)	(11.91)	(12.16)	(8.014)	(3.503)	(9.762)
Female (Yes = 1)	0.450	0	1	0.518	0.315	0.427	0.476
	(0.498)	(0)	(0)	(0.500)	(0.466)	(0.495)	(0.500)
From India (Yes = 1)	0.334	0.417	0.234	0	1	0.468	0.182
	(0.472)	(0.494)	(0.424)	(0)	(0)	(0.500)	(0.386)
N	589	324	265	392	197	314	275
Panel B: Employer							
Survey Duration (Minutes)	23.09	23.59	22.37	19.08	31.60	22.53	23.87
	(17.23)	(15.57)	(19.43)	(11.70)	(23.04)	(19.00)	(14.44)
College Education or Above	0.669	0.701	0.622	0.559	0.903	0.671	0.667
	(0.471)	(0.459)	(0.486)	(0.497)	(0.297)	(0.471)	(0.472)
Age (Employer)	34.36	32.66	36.88	35.73	31.46	27.09	44.36
	(11.02)	(9.917)	(12.07)	(11.63)	(8.962)	(3.587)	(9.906)
Female (Yes = 1)	0.404	0	1	0.485	0.232	0.341	0.490
	(0.491)	(0)	(0)	(0.500)	(0.424)	(0.475)	(0.501)
From India (Yes = 1)	0.321	0.413	0.185	0	1	0.395	0.218
	(0.467)	(0.493)	(0.389)	(0)	(0)	(0.490)	(0.414)
N	577	344	233	392	185	334	243

Notes: Standard deviations in parentheses. One observation per worker (survey 1) or employer (survey 2).

3.2 Experimental Results

A necessary prerequisite to study the source of discrimination is to find a context and a population in which discrimination occurs. Ex ante, it was not obvious that our stylized hiring experiment would satisfy this requirement. The employers knew that they were being observed as part of a research study, the stakes were low, there was no direct interaction or expectation of future interaction between the employers and the workers, and the relevant group information was represented abstractly (e.g. written text) rather than viscerally (e.g. a picture). All of these factors may attenuate the influence of animus.[11]

Despite these attenuating factors, we did find evidence of discrimination with respect to two out of three group identities: gender and nationality, but not age. Panel A of Table 6 presents the differences in average wages paid by employers to worker profiles from each group. Related to gender, male profiles are paid on average 31.90 cents, while female profiles are paid 30.85 cents, a 3.4% difference that is significant at the 1% level (p¡0.01). With respect to nationality, profiles from India are favored, earning an average of 32.85 cents, while profiles from the U.S. earn 30.71 cents, a 7.0% difference also significant at the 1% level (p¡0.01). Finally, we see modest evidence of age discrimination: subjects at or below age 33 are paid an average of 31.67 cents and those above age 33 are paid 31.14 cents, a 1.7% difference that is significant at the 5% level (p=0.02). To put these differences into context, we show a "placebo" comparison for a profile characteristic that was unlikely to be either a source of animus or a proxy for math ability: the workers preference for tea versus coffee. We find a similar level of discrimination to the level that we found for age, with tea drinkers earning a 1.7% higher average wage (31.74 vs. 31.22 cents), significant at the 5% level (p=0.03).

In Table 7, we present regression results that simultaneously control for all demographic variables, as the pairwise mean comparisons do not account for any correlations with other demographics. Columns 1-4 repeat the mean comparisons in regression format with standard errors clustered at the employer level. In column 5, we present the multivariate regression with all demographic controls and in column 6, we add employer fixed effects. In both cases, the levels of discrimination for gender and age are slightly diminished in magnitude, but remain statistically significant. To check for robustness, we run similar regressions controlling for the employer belonging to the group of interest (e.g. female) and the interaction of the two indicators to measure in-group bias (see Appendix Table 1). We find that the interaction is

[11]For example, Bar and Zussman (2019) argue that a lack of interaction may attenuate the extent of taste-based discrimination in driving test examinations.

Table 6: Wages and "Productivities", by Employee Characteristics (Hiring Task 1)

	(1)	(2)	(3)	(4)	(5)	(6)
	Group (1 or 2)		Diff.	P-Val	N(1)	N(2)
	1	2	[(1)-(2)]			
Panel A: Employers' Wage WTP, by Employee Characteristics						
Gender (1 = Male , 2 = Female)	31.90	30.85	1.05	0.00	6,306	5,234
	(12.07)	(12.23)				
Country (1 = US , 2 = India)	30.71	32.85	-2.14	0.00	7,700	3,840
	(12.20)	(11.95)				
Age (1 = Under 33 , 2 = Over 33)	31.67	31.14	0.54	0.02	6,139	5,401
	(12.00)	(12.33)				
Placebo (1 = Prefer Coffee , 2 = Prefer Tea)	31.22	31.74	-0.52	0.03	7,075	4,465
	(12.32)	(11.89)				
Panel B: Employee Productivity, by Employee Characteristics						
Gender (1 = Male , 2 = Female)	38.30	34.98	3.32	0.00	6,306	5,234
	(8.55)	(8.73)				
Country (1 = US , 2 = India)	37.01	36.36	0.65	0.00	7,700	3,840
	(8.93)	(8.49)				
Age (1 = Under 33 , 2 = Over 33)	36.96	36.60	0.37	0.03	6,139	5,401
	(8.62)	(8.98)				
Placebo (1 = Prefer Coffee , 2 = Prefer Tea)	36.64	37.03	-0.40	0.02	7,075	4,465
	(8.77)	(8.82)				

Notes: Standard deviations in parentheses. One observation per worker-employer combination. Column 4 shows the p-value from two-sample t-tests for the equality of columns 1 and 2.

insignificant for gender and marginally significant for nationality, although in the direction of favoring the out-group. For age, we find a significant interaction effect. This suggests that the null effect in Table 6 masks in-group bias by both older and younger employers.[12]

Having demonstrated moderate levels of discrimination in hiring, we turn to measured productivity differences between groups in the work task (the math quiz). The typical outcomes-based test of statistical discrimination requires mapping disparities between groups in the evaluators' relevant decision (e.g. police searches of motor vehicles) to disparities in an outcome in the evaluators' objective function (e.g. the likelihood of finding contraband).[13]

[12] Antonovics and Knight (2009) use a similar set of regressions to test for taste-based discrimination. This specification is motivated by the assumption that animus varies between groups (i.e. there is less animus toward ones in-group than out-group), but that beliefs are similar across groups (since they are taking a "standard model of statistical discrimination" as the benchmark and note that "these beliefs must be correct in equilibrium"). In Appendix Table A3, we test this assumption in our experimental environment. We find that beliefs about the gender performance gap are identical among both female and male employers. However, for nationality, we find a significant difference in beliefs, namely, Indians hold beliefs that more strongly favor the out-group (Americans).

[13] Translating the two measures may require strong modeling assumptions (e.g. whether there is hetero-

Table 7: Discrimination in Wages, by Employee Characteristics (Hiring Task 1)

	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Female	-1.05***				-0.67***	-0.80***
	(0.25)				(0.24)	(0.19)
Indian		2.14***			1.98***	2.00***
		(0.29)			(0.29)	(0.25)
Over 33			-0.54**		0.07	0.32
			(0.26)		(0.26)	(0.22)
Placebo: Prefers Tea				0.52**	0.39*	0.37**
				(0.24)	(0.24)	(0.18)
N	11,540	11,540	11,540	11,540	11,540	11,540
R^2	0.00	0.01	0.00	0.00	0.01	0.49
DepVarMean	31.90	30.71	31.67	31.22	30.18	30.18
Employer FE?	No	No	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered by employer. "DepVarMean" is the mean of the dependent variable (wage WTP) in the omitted group (e.g. Male Workers for column 1).

To simplify and make the link between the employers' objective function and decision transparent, we required employers to state a wage that is directly comparable to the outcome they are maximizing (predicting the number of correct questions by the worker) – that is, each cent for the wage is equivalent to one correct question. Therefore, we can directly compare disparities in wages to disparities in performance to measure the relative importance of (accurate) statistical versus taste-based sources. For this comparison, we will refer to both disparities as measured in "points."

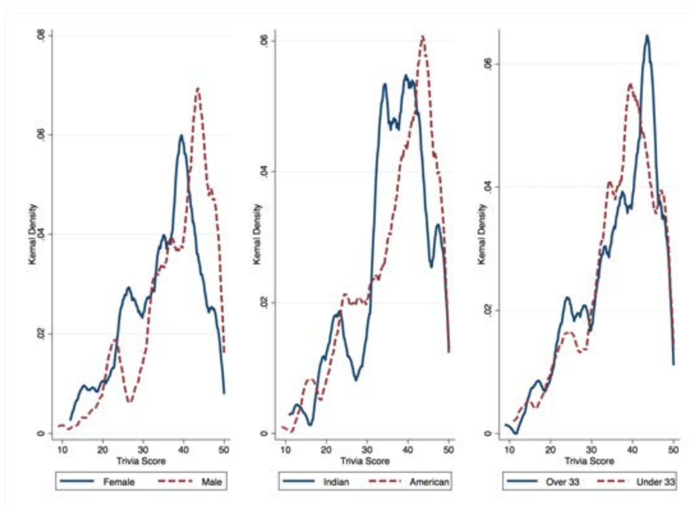
Figure 2 plots probability density functions for the number of correct answers by each sub-group. Men and subjects from India outperformed women and subjects from the U.S., respectively, whereas there was no significant difference by age. As shown in Table 6, the gap in average wages for men and women was lower than the gap in average performance (1.05 points versus 3.32 points).[14] Therefore, if we used this standard approach to separate

geneity in the search costs faced by evaluators). For discussions of these assumptions in the context of the hit-rate tests, see Antonovics and Knight 2009, Dharmapala and Ross 2004, Anwar and Fang 2006.

[14]We calculate productivity differences using the full sample of profiles observed in hiring task 1. This is a weighted sample of the original population of 577 workers (since each of the 589 employers saw independent random samples of 20 of the 577 workers). Due to the random variation in the profiles observed, the group-level averages slightly differ from those found in Table 5. For example, the male-female performance gap is 3.04 points in Table 5 and 3.32 points in this weighted sample. Note that the averages in Table 5 are the

statistical and taste-based discrimination, we would conclude that the entire 1.05 point disparity in wages is due to (accurate) statistical discrimination. Further, we would conclude that the remaining 2.17 point difference in performance suggests taste-based discrimination against men. Turning to nationality-based discrimination, there was a wage gap of -2.14 points in favor of Indians, compared to a performance gap of 0.65 points in favor of Americans. Under the standard approach, we would conclude that both the -2.14 point disparity in wages and the 0.65 point difference in performance suggest taste-based discrimination against Americans.

Figure 2: Kernal Densities of Productivities (Trivia Scores) by Group



As we argued above, this standard approach naively assumes that employers have accurate beliefs about group-level performance differences. To investigate this assumption, we elicited beliefs about performance. As a check that employers' decisions were guided by the beliefs we elicited, we correlate the wages they assigned in the prior hiring task with the predictions they made for average productivity of each group. We find modest positive correlations for all six groups of workers (Female: 0.12, Male: 0.12, India: 0.15, U.S.: 0.12, Over 33: 0.12, Under 33: 0.10). Given that we elicited beliefs after the hiring task, it is possible that part of these correlations reflect assimilation driven by cognitive dissonance (e.g. an individual first chooses to discriminate against women in setting wages and then chooses beliefs to justify this decision) or by an audience effect (e.g. an individual falsely reports beliefs that justify the discriminatory decision to the experimenter). To test for these types of cheap talk effects, we provided large incentives for accuracy for half of the employers. In Appendix

basis for the informational intervention.

Table 2, we show that beliefs are nearly identical across both incentive conditions, with none of the six comparisons being statistically different. Together these findings suggest that the employers' group-level performance predictions provide meaningful information about their beliefs.

In Table 8, we present employer beliefs about the group-level average performance (math quiz scores) that can be compared to the actual performance reported in Table 6, Panel B. Predictions about performance are lower than actual performance for all six groups. This overall underestimation is consistent with risk aversion (recall that employers face a negative payment, taken from their $0.50 bonus, if they overestimate performance). More to our point, gaps in beliefs about performance are larger than gaps in wage payments. Repeating the accounting exercise above using reported beliefs about performance, rather than actual performance (which implicitly assumes correct beliefs) leads to markedly different conclusions. By nationality, the wage gap is -2.14 points and the performance gap is 0.65 points, whereas the belief gap is -2.72 points. Thus, the *entire* wage gap is explained by inaccurate beliefs. The remaining 0.58 point (2.72 minus 2.14) difference between the belief and wage gaps suggests taste-based discrimination against Indians. This contrasts with the comparison using performance as beliefs, where we concluded that there was taste-based discrimination *in favor* of Indians. Similarly, we reach different conclusions about the sources of gender discrimination. The wage gap is 1.05 points, the performance gap is 3.32 points, and the belief gap is 1.89 points. The wage gap is still fully accounted for by beliefs, but the gap between performance and wages is partly explained by inaccurate beliefs; the difference explained by a taste-based source shrinks from 2.17 to 0.84 points. Finally, despite the minimal gap in wages and performance for age, employers believed that young workers outperform older ones. This suggests taste-based discrimination against younger workers.

The role of inaccurate beliefs in explaining part of the wage gap suggests a policy intervention: inform employers of the correct distributions and provide them with the opportunity to set wages for a different set of potential workers to see if this induces a behavioral change. To examine the effects of such a policy, we had employers assign wages to 20 profiles (hiring task 1), provided them with information about group-level math quiz score averages, and then had them evaluate 10 more profiles (hiring task 2).

Table 9 demonstrates the effects of this informational intervention. We compare the differences between the two hiring rounds (Post-Info), the differences between wages assigned to profiles of each worker demographic group (e.g. "Female"), and the resulting difference-in-differences (e.g. "Female X Post-Info"). The coefficients on "Post-Info" suggests substantial

Table 8: Beliefs about Productivity by Employee Characteristics

	(1) Group 1	(2) (1 or 2) 2	(3) Diff. [(1)-(2)]	(4) P-Val
Gender (1 = Male , 2 = Female)	34.04	32.14	1.89	0.00
	(8.26)	(8.41)		
Country (1 = US , 2 = India)	32.08	34.80	-2.72	0.00
	(8.56)	(9.44)		
Age (1 = Under 33 , 2 = Over 33)	33.41	31.57	1.84	0.00
	(8.97)	(9.00)		

Notes: Standard deviations in parentheses. One observation per employer combination. Column 4 shows the p-value from one-sample t-tests for the equality of columns 1 and 2. N = 577.

belief updating across all demographic groups, partially correcting the large level differences in hiring task 1 between wages and actual productivity (a gap of roughly 5 points on average; see Table 6). For example, Column 1 shows that the wage rate for Men increased by 1.53 points. The coefficients on the interaction between "Post-Info" and each demographic group show that the intervention had a significant effect on wage gaps by gender and nationality, but not on the more modest gap by age. The gender wage gap expands by 0.64 points, increasing to 1.71 points in hiring task 2 (Column 1). This gap remains smaller than the 3.01 performance gap between the male and female profiles. Thus the wage gap still reflects a mix of accurate statistical discrimination against women and either (a smaller amount of) taste-based discrimination against men (1.30 points instead of 2.71 points) or significant limits to belief updating in response to our intervention.[15] For nationality, the intervention

[15] Throughout the paper we find discrimination in wage rates by gender (i.e. men paid more than women). As such, we carry out the accounting exercise that would typically follow in the literature. This reveals that the gap in performance in fact exceeds that of the gap in pay. Thus, we end up concluding that there is taste-based discrimination against the group that received higher wages, e.g. men. While the literature often treats taste-based discrimination as analogous to animus or prejudice against a group, when discrimination works toward equalizing actions (e.g. wage rates), this analogy seems misplaced. For example, in many real world domains, it is illegal to discriminate on the basis of gender, religion, or race, even if those attributes correlate with the outcomes of interest. Thus, employers may fail to discriminate (or not discriminate enough, given their beliefs) against the less productive group, not because they favor them, but because they fear legal or social sanctions. Relatedly, there is often an equity-efficiency trade-off to discriminating, such that even in the absence of legal or social sanctions, an employer may wish to equalize wages across groups (for a theoretical discussion of these trade-offs in the context of racial profiling, see Durlauf 2005). Such a concern may be especially pronounced for wages, where even abstracting away from group-level attributes, there is evidence that fairness norms may contribute to observed wage compression (e.g. Breza et al. 2018).

Just as decomposing the nature of belief-based discrimination has implications for policy, the same may be true for taste-based discrimination. For example, if the basis for taste-based discrimination is animus or prejudice, then a policy that increases contact between groups may reduce disparities (Dobbie and Fryer 2015; Paluck et al. 2018). By contrast, if the behavior is instead sanction- or value-oriented, then such cut the wage gap in half, from 2.14 points down to 1.07 (Column 2). This wage gap still exceeds the performance gap of 0.42 points, again suggesting that there remains a modest amount of taste-based discrimination against Americans or limits to belief updating.

Table 9: Effect of Information: Difference-in-Differences by Hiring Task

	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Post-Info	1.53***	1.60***	1.06***	1.28***	1.94***	2.39***
	(0.29)	(0.27)	(0.31)	(0.29)	(0.44)	(0.36)
Female	-1.05***				-0.67***	-0.81***
	(0.25)				(0.24)	(0.19)
Female X Post-Info	-0.64*				-0.90**	-1.00***
	(0.37)				(0.37)	(0.29)
Indian		2.14***			1.98***	1.99***
		(0.29)			(0.29)	(0.25)
Indian X Post-Info		-1.07***			-1.20***	-1.63***
		(0.40)			(0.42)	(0.34)
Over 33			-0.54**		0.07	0.30
			(0.26)		(0.26)	(0.22)
Over 33 X Post-Info			0.41		0.14	-0.21
			(0.40)		(0.42)	(0.30)
Prefers Tea				0.52**	0.39*	0.35**
				(0.24)	(0.24)	(0.18)
Prefers Tea X Post-Info				-0.08	0.06	-0.18
				(0.38)	(0.38)	(0.27)
N	17,310	17,310	17,310	17,310	17,310	17,310
R^2	0.01	0.01	0.00	0.00	0.01	0.48
DepVarMean	31.90	30.71	31.67	31.22	30.18	30.18
Employer FE?	No	No	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered by employer. "DepVarMean" is the mean of the dependent variable (wage WTP) in the omitted group (e.g. Male Workers in Hiring Task 1 for column 1). "Post-Info" is an indicator for whether a profile came in the second hiring task (i.e. profiles 21-30 of the 30 total profiles evaluated). The observed performance (trivia score) averages for the sample of profiles observed in Hiring Task 2 are: 38.13 (Male), 35.13 (Female), 36.95 (US), 36.53 (India), 36.84 (Under 33), 36.77 (Over 33), 36.81 (Prefer Coffee), 36.79 (Prefer Tea).

---

interventions will likely have little impact. While it is difficult to imagine a simple elicitation that would allow for a parsimonious quantitative decomposition of tastes, survey measures may be able to make some headway in this endeavor. Such a decomposition is outside of the scope of this paper, but future work along these lines would enrich the model of discrimination used in economics and the tools used to identify discrimination and design policy.

Our interpretation of the second hiring task results is that employers updated their beliefs toward the signal provided by the information intervention. However, a few challenges remain in determining whether this is the causal effect of the intervention through the channel of updated beliefs. First, we did not measure beliefs again after the information intervention, and therefore, we cannot verify that employers' beliefs changed. However, as noted earlier, the fact that we find a large level effect on wages across all demographic groups is suggestive of belief updating. A second concern is that part of the change in wages could reflect an experience effect between rating the first 20 profiles and the subsequent 10 after the intervention. To investigate this channel, we perform a t-test comparing the average wages assigned in the first 10 profiles and the second 10 profiles in the first hiring task. There is no significant difference between these two sets of profiles (36.86 vs. 36.72; p=.39). A third concern is that the belief elicitation, rather than the information intervention, is what changed behavior (e.g. one possible mechanism is that stating beliefs about productivity over the given profile characteristics changes the salience of these characteristics, and therefore the weight placed on them in hiring). Since we did not have a group that had a belief elicitation but no information intervention, we cannot rule this out. Finally, the information intervention could have produced an experimenter demand effect. Subjects may have inferred that we wanted them to move their wage offers toward the performance levels, and thus done so to please the experimenter without actually changing their beliefs. While we cannot fully rule out all of these possible confounds, the information intervention serves as a proof of concept of the type of policy that could change behavior in the face of inaccurate statistical discrimination.

The results of the information intervention suggest that identifying inaccurate beliefs may have immediate policy implications for reducing discrimination. However, there are some important caveats to keep in mind when considering how this type of intervention would be implemented outside of the toy exercise in this paper. First, such an intervention may only be appropriate in contexts where the underlying target outcome (e.g. productivity) is reliably measured and reflects the appropriate counterfactual outcomes for all groups were they to be selected by the agents whose beliefs are being corrected (e.g. employers). To the first point, in some cases the underlying measures may be inaccurately recorded; for example, biased police officers are less likely to record a minority driver at a discounted speed than a white driver (Goncalves and Mello 2019). To the latter point, there are contexts in which discrimination at (often unobserved) intermediate stages renders final productivity measures unreliable due to behavioral responses. For example, minority pitchers correctly anticipate discrimination by umpires and modify their behavior, resulting in downwardly biased performance measures

(Parsons et al. 2011). Studies have also documented that bias at intermediate stages can skew final productivity measures among grocery store workers (Glover et al. 2017) and academic economists (Hengel 2019).

A second set of caveats relates to the underlying psychology of how people would respond to the information. First, selection decisions are rarely one-dimensional. Drawing attention to a (smaller than expected) productivity gap could correct beliefs while nonetheless increasing discrimination if it increases the salience of the gap and thus the weight placed on in that attribute in the selection decision. Second, as noted in footnote 1, the line between inaccurate beliefs and animus is sometimes blurry. For example, individuals may use motivated reasoning to arrive at and preserve closely held inaccurate beliefs. As noted in the footnote, we can separately identify animus-driven inaccurate beliefs through an information experiment if individuals asymmetrically update (i.e. they discard the good news about the out-group); however, such an exercise would be difficult if individuals correctly update in the short run and only exhibit motivated memory (i.e. they discard the good news with a delay, as in Zimmermann 2019). If individuals inaccurate beliefs are in fact driven by animus, then the information intervention should merely be ineffective if it is discarded. However, it is possible that individuals with such animus-driven beliefs would exhibit psychological reactance (Brehm 1966) to a perceived attempt at manipulation and thus exhibit even more biased selections as a way of expressing their agency. Third, there may be spillover effects of a social planner selectively revealing information about gaps. Consider a planner that wishes to minimize discrimination. If such a planner only reveals information about gaps when they are smaller than stereotypes would suggest, then the absence of an information intervention itself reveals information. Its not obvious how individuals would update their beliefs in response to not getting an information summary on other attributes of the individuals under consideration or how it would spillover into other selection tasks. Such updating would likely depend on agents beliefs about the threshold being used by the social planner and the motives of the planner (e.g. whether the planner is willing to allow statistical discrimination or aims to minimize all discrimination). These various concerns limit the external validity of our information intervention, but we look forward to future tests that operationalize it appropriately for field contexts.

4 Conclusion

The study of discrimination and its motives has a rich history in economics. Separating out statistical and taste-based intentions for discrimination is a useful exercise, but as our literature review illustrates, is one that has typically relied heavily on the assumption of accurate beliefs. However, there are many reasons to suspect that beliefs may not always be accurate (both psychological explanations of bias and also due to limited information of rational agents). In this paper, we document and provide an example for the misinterpretation of data that can result from researchers assuming accurate beliefs.

Going forward, we argue that researchers studying motives of discrimination should in all situations consider and discuss the possibility of inaccurate beliefs and elicit beliefs when possible. By doing so, researchers are less likely to mistakenly assign motives of animus or accurate statistical discrimination when those may actually not be the actual intentions. Further, this research speaks to the need for continued work such as Bordalo et al. 2016 that might help identify situations when inaccurate beliefs are especially likely to be prevalent. As research begins to identify situations where inaccurate beliefs are a driving factor for discrimination, future work will hopefully also begin to develop policy interventions that are able to effectively improve beliefs and thereby reduce discrimination.

References

Aigner, D. J. and G. G. Cain (1977): "Statistical Theories of Discrimination in Labor Markets," *ILR Review*, 30, 175–187.

Antonovics, K. and B. G. Knight (2009): "A New Look at Racial Profiling: Evidence from the Boston Police Department," *Review of Economics and Statistics*, 91, 163–177.

Anwar, S. and H. Fang (2006): "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence," *American Economic Review*, 96, 127–151.

Arnold, D., W. Dobbie, and C. Yang (2018): "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, 1885–1932.

Arrow, K. J. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter and A. Rees, Princeton, NJ: Princeton University Press.

Bar, R. and A. Zussman (2019): "Identity and Bias: Insights from Driving Tests," *Working Paper*, 1–45.

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

Bertrand, M. and E. Duflo (2017): "Field Experiments on Discrimination," *Handbook of Field Experiments*, 1, 110.

Bohren, J. A., A. Imas, and M. Rosenberg (2019): "The Dynamics of Discrimination: Theory and Evidence," *Working Paper*.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016): "Stereotypes," *Quarterly Journal of Economics*, 1753–1794.

Brehm, J. W. (1966): *A Theory of Psychological Reactance*, New York: Academic Press.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The morale effects of pay inequality," *Quarterly Journal of Economics*, 133, 611–663.

Cameron, S. V. and J. J. Heckman (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109, 455–499.

Charles, K. K. and J. Guryan (2011): "Studying Discrimination: Fundamental Challenges and Recent Progress," *Annual Review of Economics*, 3, 479–511.

Dharmapala, D. and S. L. Ross (2004): "Racial bias in motor vehicle searches: Additional theory and evidence," *Contributions to Economic Analysis and Policy*, 3, 89–111.

Dobbie, W. and R. G. Fryer (2015): "The Impact of Youth Service on Future Outcomes: Evidence from Teach for America," *The B.E. Journal of Economic Analysis and Policy*, 15, 1031–1066.

Durlauf, S. N. (2005): "Racial Profiling as a Public Policy Question: Efficiency, Equity, and Ambiguity," *American Economic Review*, 95, 132–136.

Fershtman, C. and U. Gneezy (2001): "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, February, 351–377.

Glover, D., A. Pallais, and W. Pariente (2017): "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *Quarterly Journal of Economics*, 1219–1260.

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance in competitive Environments: Gender differences," *Quarterly Journal of Economics*, 1049–1074.

Goncalves, F. and S. Mello (2019): "A Few Bad Apples? Racial Bias in Policing," *Working Paper*, 1–79.

Hedegaard, M. S. and J.-R. Tyran (2018): "The Price of Prejudice," *American Economic Journal: Applied Economics*, 10, 40–63.

Hengel, E. (2019): "Publishing while female," *Working Paper*, 1–67.

Hilton, J. L. and W. V. Hippel (1996): "Stereotypes," *Annual Review of Psychology*, 47, 237–271.

Jensen, R. (2010): "The (perceived) returns to education and the demand for schooling," *Quarterly Journal of Economics*, 515–548.

Judd, C. M. and B. Park (1993): "Definition and assessment of accuracy in social stereotypes," *Psychological Review*, 100, 109–128.

Kravitz, D. A. and J. Platania (1993): "Attitudes and Beliefs About Affirmative Action: Effects of Target and of Respondent Sex and Ethnicity," *Journal of Applied Psychology*, 78, 928–938.

List, J. A. (2004): "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics*, 49–89.

Mobius, M. and T. Rosenblat (2006): "Why Beauty Matters," *American Economic Review*, 96, 222–235.

Norton, M. I. and D. Ariely (2011): "Building a better America-one wealth quintile at a time," *Perspectives on Psychological Science*, 6, 9–12.

Paluck, E. L., S. Green, and D. P. Green (2018): "The Contact Hypothesis Reevaluated," *Behavioural Public Policy*, 1–30.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011): "Strike Three: Discrimination, Incentives, and Evaluation," *American Economic Review*, 101, 1410–1435.

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659–661.

Schneider, D., A. Hastorf, and P. Ellsworth (1979): *Person Perception*, Reading, MA: Addison-Wesley.

Zimmermann, F. (2019): "The Dynamics of Motivated Beliefs," *Working Paper*.

A Appendix: Additional Tables

Table A1: Effects of Large Incentives for Accurate Predictions

	(1)	(2)	(3)	(4)
	Incentivized?		Diff.	P-Val
	No	Yes	[(1)-(2)]	
Prediction for Female Workers	32.36	31.93	0.44	0.53
	(7.71)	(9.08)		
Prediction for Male Workers	34.22	33.86	0.36	0.60
	(7.37)	(9.08)		
Prediction for Indian Workers	35.29	34.31	0.98	0.21
	(8.49)	(10.30)		
Prediction for US Workers	32.28	31.87	0.41	0.56
	(8.21)	(8.90)		
Prediction for Over 33 Workers	31.95	31.19	0.75	0.32
	(8.39)	(9.58)		
Prediction for Under 33 Workers	33.73	33.09	0.64	0.39
	(8.58)	(9.35)		
Observations	290	287		

Notes: Standard deviations in parentheses. One observation per employer. The joint f-statistic from regression of an indicator for the "Incentivized" treatment on set of employer observables characteristics in Table 1, Panel B (duration, college education or above, age, female, from India) is 1.25 (p=0.2863).

Table A2: In-Group Bias Test (Hiring Task 1)

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Female Worker	-1.42***			-1.20***
	(0.27)			(0.27)
Female Employer	1.78***			1.91***
	(0.69)			(0.72)
Female Worker X Employer	0.26			0.41
	(0.43)			(0.42)
Indian Worker		2.04***		1.88***
		(0.31)		(0.31)
Indian Employer		0.99		1.70**
		(0.70)		(0.74)
Indian Worker X Employer		-0.79*		-0.82*
		(0.48)		(0.48)
Over 33 Worker			-0.86***	-0.39
			(0.29)	(0.28)
Over 33 Employer			0.31	0.22
			(0.69)	(0.71)
Over 33 Worker X Employer			1.10***	1.19***
			(0.42)	(0.41)
N	17,310	17,310	17,310	17,310
R^2	0.01	0.01	0.00	0.02
DepVarMean	31.90	30.71	31.67	31.67

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors clustered by employer. "DepVarMean" is the mean of the dependent variable (wage WTP) in the omitted group (e.g. Male Workers evaluated by Male Employers for column 1).

Table A3: In-Group vs. Out-Group Beliefs about Productivity by Employee Characteristiscs

	(1)	(2)	(3)	(4)	(5)	(6)
	Out vs. In Group		Diff.	P-Val	N(1)	N(2)
	Out	In	[(1)-(2)]			
Prediction for Female Workers	31.70	32.79	-1.09	0.13	344	233
	(8.78)	(7.81)				
Prediction for Male Workers	34.68	33.60	1.09	0.12	233	344
	(6.59)	(9.20)				
Prediction for Indian Workers	36.09	32.06	4.04	0.00	392	185
	(7.10)	(12.67)				
Prediction for US Workers	30.46	32.84	-2.38	0.00	185	392
	(12.04)	(6.15)				
Prediction for Over 33 Workers	30.92	32.47	-1.55	0.04	334	243
	(9.82)	(7.66)				
Prediction for Under 33 Workers	33.85	33.09	0.77	0.31	243	334
	(7.03)	(10.14)				

Notes: Standard deviations in parentheses. "In-Group" refers to a match in the characteristic between the employer and the group of workers over which they are making a prediction, e.g. column1, row 1 is the average prediction made by female employers about the average productivity of female workers.

B Appendix: Papers Included in Literature Survey

In this section, we list the citation for each paper included in the literature survey.

Abrevaya, J. and D. S. Hamermesh (2012): "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?" *The Review of Economics and Statistics*, 94, 202–207.

Acemoglu, D. and J. Angrist (2001): "Consequences of Employment Protection? The Case of the Americans with Disabilities Act," *Journal of Political Economy*, 109, 915–957.

Agan, A. and S. Starr (2017): "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *The Quarterly Journal of Economics*, 133, 191–235.

Alan, S., S. Ertac, and I. Mumcu (2018): "Gender Stereotypes in the Classroom and Effects on Achievement," *The Review of Economics and Statistics*, 100, 876–890.

Alesina, A. and E. L. Ferrara (2014): "A Test of Racial Bias in Capital Sentencing," *The American Economic Review*, 104, 3397–3433.

Altonji, J. G. and C. R. Pierret (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

Antecol, H. and P. Kuhn (2000): "Gender as an Impediment to Labor Market Success: Why Do Young Women Report Greater Harm?" *Journal of Labor Economics*, 18, 702–728.

Antonovics, K. and B. G. Knight (2009): "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics*, 91, 163–177.

Anwar, S., P. Bayer, and R. Hjalmarsson (2012): "The Impact of Jury Race in Criminal Trials," *The Quarterly Journal of Economics*, 127, 1017–1055.

Anwar, S. and H. Fang (2006): "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *The American Economic Review*, 96, 127–151.

Arai, M. and P. SkogmanThoursie (2009): "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings," *Journal of Labor Economics*, 27, 127–147.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial Bias in Bail Decisions," *The Quarterly Journal of Economics*, 133, 1885–1932.

Aslund, O., L. Hensvik, and O. N. Skans (2014): "Seeking Similarity: How Immigrants and Natives Manage in the Labor Market," *Journal of Labor Economics*, 32, 405–441.

Ayres, I. and P. Siegelman (1995): "Race and Gender Discrimination in Bargaining for a New Car," *The American Economic Review*, 85, 304–321.

Bagues, M. F. and B. Esteve-Volart (2010): "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *The Review of Economic Studies*, 77, 1301–1328.

Baldwin, M. and W. G. Johnson (1992): "Estimating the Employment Effects of Wage Discrimination," *The Review of Economics and Statistics*, 74, 446–455.

Bar, R. and A. Zussman (2017): "Customer Discrimination: Evidence from Israel," *Journal of Labor Economics*, 35, 1031–1059.

Barcellos, S. H., L. S. Carvalho, and A. Lleras-Muney (2014): "Child Gender and Parental Investments In India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics*, 6, 157–189.

Barto, V., M. Bauer, J. Chytilov, and F. Matjka (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *The American Economic Review*, 106, 1437–1475.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009): "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics*, 124, 1497–1540.

Beaman, L., N. Keleher, and J. Magruder (2018): "Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi," *Journal of Labor Economics*, 36, 121–157.

Berkovec, J. A., G. B. Canner, S. A. Gabriel, and T. H. Hannan (1998): "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending," *The Review of Economics and Statistics*, 80, 241–250.

Bertrand, M. and S. Mullainathan (2004): "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*, 94, 991–1013.

Biddle, J. E. and D. S. Hamermesh (1998): "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics*, 16, 172–201.

Black, S. E. and P. E. Strahan (2001): "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry," *The American Economic Review*, 91, 814–831.

Blanchflower, D. G., P. B. Levine, and D. J. Zimmerman (2003): "Discrimination in the Small-Business Credit Market," *The Review of Economics and Statistics*, 85, 930–943.

Bollinger, C. R. (2003): "Measurement Error in Human Capital and the Black-White Wage Gap," *The Review of Economics and Statistics*, 85, 578–585.

Botelho, F., R. A. Madeira, and M. A. Rangel (2015): "Racial Discrimination in Grading: Evidence from Brazil," *American Economic Journal: Applied Economics*, 7, 37–52.

Breda, T. and S. T. Ly (2015): "Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France," *American Economic Journal: Applied Economics*, 7, 53–75.

Brevoort, K. P. (2011): "Credit Card Redlining Revisited," *The Review of Economics and Statistics*, 93, 714–724.

Brueckner, J. and Y. Zenou (2003): "Space and Unemployment: The LaborMarket Effects of Spatial Mismatch," *Journal of Labor Economics*, 21, 242–262.

Burgess, S. and E. Greaves (2013): "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities," *Journal of Labor Economics*, 31, 535–576.

Butcher, K. F., K. H. Park, and A. M. Piehl (2017): "Comparing Apples to Oranges: Differences in Womens and Mens Incarceration and Sentencing Outcomes," *Journal of Labor Economics*, 35, S201–S234.

Carruthers, C. K. and M. H. Wanamaker (2017): "Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap," *Journal of Labor Economics*, 35, 655–696.

Case, A. and C. Paxson (2008): "Stature and Status: Height, Ability, and Labor Market Outcomes," *Journal of Political Economy*, 116, 499–532.

Charles, K. and J. Guryan (2008): "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination," *Journal of Political Economy*, 116, 773–809.

Combes, P.-P., B. Decreuse, M. Laounan, and A. Trannoy (2016): "Customer Discrimination and Employment Outcomes: Theory and Evidence from the French Labor Market," *Journal of Labor Economics*, 34, 107–160.

Dahl, G. B. and E. Moretti (2008): "The Demand for Sons," *The Review of Economic Studies*, 75, 1085–1120.

Donald, S. G. and D. S. Hamermesh (2006): "What Is Discrimination? Gender in the American Economic Association, 1935-2004," *The American Economic Review*, 96, 1283–1292.

Eckstein, Z. and K. I. Wolpin (1999): "Estimating the Effect of Racial Discrimination on First Job Wage Offers," *The Review of Economics and Statistics*, 81, 384–392.

Edelman, B., M. Luca, and D. Svirsky (2017): "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, 9, 1–22.

Elliehausen, G. E. and E. C. Lawrence (1990): "Discrimination in Consumer Lending," *The Review of Economics and Statistics*, 72, 156–160.

Ewens, M., B. Tomlin, and L. C. Wang (2014): "Statistical Discrimination or Prejudice? A Large Sample Field Experiment," *The Review of Economics and Statistics*, 96, 119–134.

Fershtman, C. and U. Gneezy (2001): "Discrimination in a Segmented Society: An Experimental Approach," *The Quarterly Journal of Economics*, 116, 351–377.

Foote, C., W. Whatley, and G. Wright (2003): "Arbitraging a Discriminatory Labor Market: Black Workers at the Ford Motor Company, 19181947," *Journal of Labor Economics*, 21, 493–532.

Foster, A. D. and M. R. Rosenzweig (1996): "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market," *The Review of Economic Studies*, 63, 347–374.

Fryer, R. G. and S. D. Levitt (2010): "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, 2, 210–240.

Gardeazabal, J. and A. Ugidos (2004): "More on Identification in Detailed Wage Decompositions," *The Review of Economics and Statistics*, 86, 1034–1036.

Gayle, G.-L. and L. Golan (2012): "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968-1997," *The Review of Economic Studies*, 79, 227–267.

Glover, D., A. Pallais, and W. Pariente (2017): "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *The Quarterly Journal of Economics*, 132, 1219–1260.

Goldin, C. and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *The American Economic Review*, 90, 715–741.

Goldsmith, A. H., D. Hamilton, and W. Darity (2006): "Shades of Discrimination: Skin Tone and Wages," *The American Economic Review*, 96, 242–245.

Gong, J., Y. Lu, and H. Song (2018): "The Effect of Teacher Gender on Students Academic and Noncognitive Outcomes," *Journal of Labor Economics*, 36, 743–778.

Hamermesh, D. S. and J. E. Biddle (1994): "Beauty and the Labor Market," *The American Economic Review*, 84, 1174–1194.

Hanna, R. N. and L. L. Linden (2012): "Discrimination in Grading," *American Economic Journal: Economic Policy*, 4, 146–168.

Heap, S. P. H. and D. J. Zizzo (2009): "The Value of Groups," *The American Economic Review*, 99, 295–323.

Hedegaard, M. S. and J.-R. Tyran (2018): "The Price of Prejudice," *American Economic Journal: Applied Economics*, 10, 40–63.

Hendricks, W., L. DeBrock, and R. Koenker (2003): "Uncertainty, Hiring, and Subsequent Performance: The NFL Draft," *Journal of Labor Economics*, 21, 857–886.

Hersch, J. (2008): "Profiling the New Immigrant Worker: The Effects of Skin Color and Height," *Journal of Labor Economics*, 26, 345–386.

Heywood, J. S. and D. Parent (2012): "Performance Pay and the White-Black Wage Gap," *Journal of Labor Economics*, 30, 249–290.

Hirsch, B. and D. Macpherson (2004): "Wages, Sorting on Skill, and the Racial Composition of Jobs," *Journal of Labor Economics*, 22, 189–210.

Hirsch, B., T. Schank, and C. Schnabel (2010): "Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked EmployerEmployee Data from Germany," *Journal of Labor Economics*, 28, 291–330.

Hjort, J. (2014): "Ethnic Divisions and Production in Firms," *The Quarterly Journal of Economics*, 129, 1899–1946.

Holzer, H. J. and K. R. Ihlanfeldt (1998): "Customer Discrimination and Employment Outcomes for Minority Workers," *The Quarterly Journal of Economics*, 113, 835–867.

Ichino, A. and E. Moretti (2009): "Biological Gender Differences, Absenteeism, and the Earnings Gap," *American Economic Journal: Applied Economics*, 1, 183–218.

Ihlanfeldt, K. R. and M. V. Young (1994): "Intrametropolitan Variation in Wage Rates: The Case of Atlanta Fast-Food Restaurant Workers," *The Review of Economics and Statistics*, 76, 425–433.

Jayachandran, S. and I. Kuziemko (2011): "Why do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India," *The Quarterly Journal of Economics*, 126, 1485–1538.

Kelchtermans, S. and R. Veugelers (2013): "Top Research Productivity and its Persistence: Gender as a Double-Edged Sword," *The Review of Economics and Statistics*, 95, 273–285.

Kenney, G. M. and D. A. Wissoker (1994): "An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers," *The American Economic Review*, 84, 674–683.

Knepper, M. (2018): "When the Shadow Is the Substance: Judge Gender and the Outcomes of Workplace Sex Discrimination Cases," *Journal of Labor Economics*, 36, 623–664.

Knowles, J., N. Persico, and P. Todd (2001): "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 109, 203–229.

Kreisman, D. and M. A. Rangel (2015): "On the Blurring of the Color Line: Wages and Employment for Black Males of Different Skin Tones," *The Review of Economics and Statistics*, 97, 1–13.

Kuhn, P. and K. Shen (2013): "Gender Discrimination in Job Ads: Evidence from China," *The Quarterly Journal of Economics*, 128, 287–336.

Lang, K. and M. Manove (2011): "Education and Labor Market Discrimination," *The American Economic Review*, 101, 1467–1496.

Lange, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

Leonard, J. S., D. I. Levine, and L. Giuliano (2010): "Customer Discrimination," *The Review of Economics and Statistics*, 92, 670–678.

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

——— (2006): ""Friend or Foe?" A Natural Experiment of the Prisoner's Dilemma," *The Review of Economics and Statistics*, 88, 463–471.

Mechtenberg, L. (2009): "Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages," *The Review of Economic Studies*, 76, 1431–1459.

Miller, A. R. and C. Segal (2012): "Does Temporary Affirmative Action Produce Persistent Effects? A Study of Black and Female Employment in Law Enforcement," *The Review of Economics and Statistics*, 94, 1107–1125.

Mobius, M. M. and T. S. Rosenblat (2006): "Why Beauty Matters," *The American Economic Review*, 96, 222–235.

Nardinelli, C. and C. Simon (1990): "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball," *The Quarterly Journal of Economics*, 105, 575–595.

Neal, D. A. and W. R. Johnson (1996): "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104, 869–895.

Neggers, Y. (2018): "Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India," *American Economic Review*, 108, 1288–1321.

Neumark, D., R. J. Bank, and K. D. V. Nort (1996): "Sex Discrimination in Restaurant Hiring: An Audit Study," *The Quarterly Journal of Economics*, 111, 915–941.

Neumark, D. and W. A. Stock (1999): "Age Discrimination Laws and Labor Market Efficiency," *Journal of Political Economy*, 107, 1081–1125.

Oettinger, G. S. (1996): "Statistical Discrimination and the Early Career Evolution of the Black- White Wage Gap," *Journal of Labor Economics*, 14, 52–78.

Ondrich, J., S. Ross, and J. Yinger (2003): "Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses from Black Customers?" *The Review of Economics and Statistics*, 85, 854–873.

Oreopoulos, P. (2011): "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes," *American Economic Journal: Economic Policy*, 3, 148–171.

Park, K. H. (2017): "Do Judges Have Tastes for Discrimination? Evidence from Criminal Courts," *The Review of Economics and Statistics*, 99, 810–823.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011): "Strike Three: Discrimination, Incentives, and Evaluation," *The American Economic Review*, 101, 1410–1435.

Persico, N., A. Postlewaite, and D. Silverman (2004): "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height," *Journal of Political Economy*, 112, 1019–1053.

Plug, E., D. Webbink, and N. Martin (2014): "Sexual Orientation, Prejudice, and Segregation," *Journal of Labor Economics*, 32, 123–159.

Price, J. and J. Wolfers (2010): "Racial Discrimination Among NBA Referees," *The Quarterly Journal of Economics*, 125, 1859–1887.

Rapaport, C. (1995): "Apparent Wage Discrimination when Wages are Determined by Nondiscriminatory Contracts," *The American Economic Review*, 85, 1263–1277.

Rehavi, M. M. and S. B. Starr (2014): "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy*, 122, 1320–1354.

Ritter, J. A. and L. J. Taylor (2011): "Racial Disparity in Unemployment," *The Review of Economics and Statistics*, 93, 30–42.

Rubinstein, Y. and D. Brenner (2014): "Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labour Market Discrimination," *The Review of Economic Studies*, 81, 389–425.

Shayo, M. and A. Zussman (2011): "Judicial Ingroup Bias in the Shadow of Terrorism," *The Quarterly Journal of Economics*, 126, 1447–1484.

——— (2017): "Conflict and the Persistence of Ethnic Bias," *American Economic Journal: Applied Economics*, 9, 137–65.

Szymanski, S. (2000): "A Market Test for Discrimination in the English Professional Soccer Leagues," *Journal of Political Economy*, 108, 590–603.

Tootell, G. M. B. (1996): "Redlining in Boston: Do Mortgage Lenders Discriminate Against Neighborhoods?" *The Quarterly Journal of Economics*, 111, 1049–1079.

Weber, A. and C. Zulehner (2014): "Competition and Gender Prejudice: Are Discriminatory Employers Doomed to Fail?" *Journal of the European Economic Association*, 12, 492–521.

Wolfers, J. (2006): "Diagnosing Discrimination: Stock Returns and Ceo Gender," *Journal of the European Economic Association*, 4, 531–541.

Wozniak, A. (2015): "Discrimination and the Effects of Drug Testing on Black Employment," *The Review of Economics and Statistics*, 97, 548–566.

C Appendix: Qualtrics Surveys

In this section, we include the Qualtrics survey used in the MTurk worker math trivia task, followed by the survey used in the MTurk employer hiring task. Survey block titles (not shown to participants) are in bold and underlined.

CHICAGO BOOTH

The University of Chicago Booth School of Business

Intro

Thank you for participating in this survey!

The survey has two parts. In the first part you will answer some very basic demographics questions. In the second part you will answer 50 multiple-choice math questions.

We are interested in determining how many of these math questions you can get right without any help. So please **do not** use a calculator or look up the answers online, but rather just do your best. The number of questions you answer correctly will not affect your payment in any way.

Demographics

Please answer the personal profile questions below:

What is your favorite color?

What is your favorite movie?

Do you prefer coffee or tea?

- ○ Tea

○ Coffee

What is your age?

What is your gender?

○ Female

○ Male

What is your favorite subject in high school?

What is your favorite sport?

Math

Q1. What is the square root of 289?

○ 17

○ 19

○ 15

○ 21

Q2. 4-8*9/2 = ?

○ -6

○ -32

○ -18

○ -12

Q3. 3^5 = ?

○ 243

○ 405

- 729
- 81

Q4. 5*6*7 = ?

- 233
- 210
- 240
- 180

Q5. What is the reduced form of the fraction 70/42?

- 7/5
- 14/10
- 5/3
- 10/6

Q6. What is the cubic root of 64?

- 4
- 6
- 5
- 3

Q7. (4+5)/5 = ?

- 6.25
- 1
- 1.8
- 5

Q8. x+2 < 18/3. Which of the following is necessarily **false**?

- x > 4
- x < 3
- x > 3
- x < 4

Q9. x^5 * x^8 = ?

- x^11
- x^14
- x^13
- x^12

Q10. Which of the following is approximately equal to 0.833?

- 5/6
- 4/5
- 6/7
- 3/4

Q11. x=5, y=6, z=7, then what is xy/(z-4)?

- 8
- 10
- 6
- 4

Q12. Which of the following is the closest integer to 45/7?

- 6
- 5
- 7
- 8

Q13. Which of the following is an integer multiple of 9?

- 3618
- 3619
- 3617
- 3620

Q14. 10/5+34-4 = ?

- 32
- 34
- 30
- 36

Q15. (x-1)*(x^2-4)=0, then which of the following **cannot** be x?

- 2
- -2
- -1
- 1

Q16. What is the square root of 196?

- 12
- 13
- 15
- 14

Q17. 5-6/(18/9) = ?

- 2
- -2
- -0.5
- 0.5

Q18. (y+9)*(y^2-121)=0, then which of the following **cannot** be y?

- 11
- 9
- -9
- -11

Q19. Which of the following is an integer multiple of 11?

- 133
- 130
- 132
- 131

Q20. 5+6+7+8+9+10 = ?

- 45
- 51
- 42
- 48

Q21. What is the binary form of 7?

- 101
- 100
- 111
- 110

Q22. 35/7+1 = ?

- 6
- 4
- 7
- 5

Q23. 24/4/3 = ?

- 4
- 3
- 1
- 2

Q24. Which of the following in an integer multiple of 4?

- 66
- 62

- 56
- 74

Q25. Which of the following is **not** a prime number?

- 4
- 2
- 3
- 5

Q26. 2*3*4*5 = ?

- 720
- 24
- 240
- 120

Q27. 6^3=?

- 216
- 432
- 36
- 128

Q28. (4*2+7*8)/4=?

- 20
- 24
- 16
- 12

Q29. Which of the following is a prime number?

- 23
- 27
- 21

- 25

Q30. 16 < x+8 < 26. Which of the following could x be?

- 23
- 18
- 13
- 8

Q31. 45+3-1 = ?

- 48
- 46
- 47
- 49

Q32. x^6 + x^6 = ?

- x^12
- x^36
- (2x)^6
- 2x^6

Q33. Which of the following fractions cannot be further reduced?

- 7/35
- 46/2
- 3/5
- 3/6

Q34. Which of the following numbers has an integer square root?

- 40
- 48
- 32
- 36

Q35. 5*(7+3)+5-4 = ?

- 51
- 55
- 39
- 32

Q36. Which of the following is **not** a factor of 30?

- 3
- 5
- 2
- 4

Q37. x^6 / x^4 = ?

- x^24
- x^10
- x^2
- x^(2/3)

Q38. 56/8 = ?

- 6
- 5
- 7
- 8

Q39. 2^4 - 3^3 = ?

- 11
- 9
- -11
- -9

Q40. (18+19+20)/3 = ?

- 20
- 21
- 19
- 18

Q41. Twenty **cannot** be divided by which of the following?

- 5
- 3
- 2
- 4

Q42. 4+8+12+16 = ?

- 40
- 20
- 25
- 45

Q43. (x^5)^3 = ?

- 5x^3
- 3x^5
- x^15
- x^8

Q44. Which of the following is the correct factorization of 36?

- 4 * 9
- 2^2 * 3^2
- 4 * 3^2
- 2^2 * 9

Q45. 3^2*2 = ?

- 18
- 42

- 81
- 24

Q46. -2*(-3-8) = ?

- -14
- 14
- 22
- -22

Q47. Which of the following is an integer multiple of 5?

- 44
- 46
- 43
- 45

Q48. x^4 = 81. What is x?

- 9
- 20.5
- 3
- 6

Q49. 76/4 = ?

- 18
- 19
- 17
- 20

Q50. Which of the following is negative?

- 2^2
- (-2)^2
- (-2)^3

- 2^3

Final

Thank you for your participation. In addition to your base payment, we may put a small bonus into your account sometime in the next few weeks. Who receives the bonus payment is determined by a different experiment that we are doing and is unrelated to how well you did in the task. Please just think of it as an additional appreciation for your efforts.

CHICAGO BOOTH

The University of Chicago Booth School of Business

Introduction

Thank you for participating in this survey.

The survey has four parts. You will first answer some simple demographic questions. Then you will answer three sets of questions related to people's performance in math questions.

The survey will take approximately 20 minutes.

Please enter your M-Turk ID:

What is your gender?

- Male
- Female

What is your age?

Please indicate the highest level of education you have completed.

- Less than High School
- High School or equivalent
- Vocational/Technical School (2 year)

- Some College
- College Graduate (4 year)
- Master's Degree (MS)
- Doctoral Degree (PhD)
- Professional Degree (MD, JD, etc.)
- Other

Setup

We recently paid many people to answer 50 math questions each. Here are some examples of the types of math questions we asked:

Question 1: What is the square root of 289?
    Choices: 15, 17, 19, 21

Question 2: 4-8*9/2 = ?
    Choices: -6, -12, -18, -32

Question 3: What is the reduced form of the fraction 70/42?
    Choices: 5/3, 10/6, 7/5, 14/10

Question 4: x^5 * x^8 = ?
    Choices: x^11, x^12, x^13, x^14

Question 5: What is the binary form of 7?
    Choices: 100, 101, 110, 111

On average, participants answered 36.95 out of 50 questions correctly.

Today, you are going to be an employer. You will hire one of the people who answered our math questions. The person you hire will be given a bonus (the wage that you choose to pay them) and in return you will receive money based on how many of the math questions they answered correctly.

Specifically, we are going to provide you with the profiles of 20 people (potential

Specifically, we are going to provide you with the profiles of 20 people (potential employees) who answered our math questions. For each of the 20 people that we present, you will indicate what is the highest wage (between 0 and 50 cents) you would be willing to pay that person. In return, you will be paid 1 cent for every question that the person you end up hiring answered correctly.

After you indicate the highest wage you would be willing to give to each employee, we will randomly draw a number between 0 and 50. If the wage you chose for the employee is equal to or higher than the randoly-drawn number, then that employee will receive the random number as a bonus, and you will receive a profit equal to the number of correct answers given by the individual minus the random number that was drawn. If the highest wage you were willing to pay the individual is lower than the random number, you will not hire the employee and neither you nor the employee will receive a bonus.

Let's walk through an example of how this works. Below is an example of a potential employee profile that you might see:

Country:	United States
Gender:	Female
Age:	63
Favorite High School Subject:	English
Favorite Sport:	Gymnastics
Favorite Color:	Sea Green
Favorite Movie:	Overboard
Prefers Coffee/Tea:	Tea

We will ask you the highest amount you would be willing to pay this employee. Let's imagine that you say you would be willing to pay this employee 40 cents.

We will then select a random number between 0 and 50. Let's say the randomly-selected number is 20. Because the highest wage you are willing to pay that person is more than 20, you will "hire" this person and they will receive 20 cents. You will then be paid based on the number of correct answers this person gave. If the person answered 30 questions correctly, you will be paid 10 cents (30-20). If the person answered 10 questions correctly, you will be paid -10 cents (10-20).

Imagine instead that the randomly-drawn number is 45. Then you will not "hire" the person and neither you nor the person will receive a bonus.

In today's task, you will actually only hire 1 person. After you decide the most you would be willing to pay to each of the 20 people we present, we will randomly select one profile to use as the actual hiring decision. We will then draw the random number between 0 and 50 and pay you the profit you've earned for that profile and pay the wage to the person whose profile you pick. We are going to automatically give you a $0.50 bonus in addition to what money you make with your hiring decision (so that there is no way you end up owing us any money after doing this task).

Just to make sure you understand, imagine you saw a profile and entered **43** as the highest amount you would be willing to pay. Now imagine the random number generated was **18** and the individual answered **10** questions correctly.

How many cents would you have to pay the individual?

How many cents would you be paid based on the individual's performance (before subtracting the wage you have to pay the individual)?

Suppose instead that you had reported **15** as the highest wage you would pay, and everything else stayed the same:

How many cents would you have to pay the individual?

How many cents would you be paid based on the individual's performance (before subtracting the wage you have to pay the individual)?

Hidden Generator

Required

You have completed ${lm://Field/1} of 20 required profiles.

Please indicate the **highest wage** you would be willing to pay this employee in the text box below.

Enter the highest wage you would be willing to pay this individual (between 0 and 50 cents):

Prediction

Thank you for completing part 2 of 4 of this survey. As promised, we will randomly select one profile and pay you your $0.50 bonus plus whatever money you make based on the hiring of the randomly-selected profile.

For the third part of this survey, please answer the six questions below. Please remember that people answered **36.95** questions correctly on average.

On average, how many math questions out of 50 do you think **women** answered correctly?

On average, how many math questions out of 50 do you think **men** answered correctly?

On average, how many math questions out of 50 do you think **people from the United States** answered correctly?

On average, how many math questions out of 50 do you think **people from India** answered correctly?

On average, how many math questions out of 50 do you think **people below or at the age of 33** answered correctly?

On average, how many math questions out of 50 do you think **people above the age of 33** answered correctly?

Thank you for completing part 2 of 4 of this survey. As promised, we will randomly select one profile and pay you your $0.50 bonus plus whatever money you make based on the hiring of the randomly-selected profile.

For the third part of this survey, please answer the six questions below. Please remember that people answered **36.95** questions correctly on average.

You have the chance to earn a significant bonus if you answer these questions correctly. We will randomly pick one question and pay you $5 minus your deviation from the correct answer. For example, if your answer for the randomly picked question is 40 and the truth is 37, then you will get a $2 bonus. You cannot receive a negative bonus. So, please answer the questions as carefully as possible so that you can potentially win a large bonus.

On average, how many math questions out of 50 do you think **women** answered correctly?

On average, how many math questions out of 50 do you think **men** answered correctly?

On average, how many math questions out of 50 do you think **people from the United States** answered correctly?

On average, how many math questions out of 50 do you think **people from India** answered correctly?

On average, how many math questions out of 50 do you think **people below or at the age of 33** answered correctly?

On average, how many math questions out of 50 do you think **people above the age of 33** answered correctly?

Truth

Here are the correct answers for the 6 questions you have answered above. On average:

- Women got **35.28** questions right.
- Men got **38.32** questions right.
- People from the U.S. got **37.14** questions right.
- People from India got **36.58** questions right.
- People below or at the age of 33 got **37.10** questions right.
- People above the age of 33 got **36.79** questions right.

Now that you have learned those facts, we would like you to work on 10 more profiles.

As before, after you finish working on those 10 additional profiles, we will randomly select one profile and randomly select a number between 0 and 50. If your highest wage is more than the randomly-selected number, we will pay you the profit you've earned for that profile as a bonus and pay the wage to the person who answered the math questions.

Extra

You have completed ${lm://Field/1} of 10 additional profiles.

Please indicate the **highest wage** you would be willing to pay this employee in the text box below.

Enter the highest wage you would be willing to pay this individual (between 0 and 50 cents):

Final

Thank you for your participation. We will calculate your bonus based on the rules specified in each part above, and pay the bonus to your account within a week.

If you have any additional comments about this survey, please provide them below. (Optional)