

The (in)visible hand: role of race and sex in job selection decisions

PHILIPP DOERRENBERG* DENVIL DUNCAN[†] DANYANG LI[‡]

January 16, 2020

Preliminary - COMMENTS WELCOMED - Do Not Cite

Abstract

We estimate the impact of employers' sex and race on the labor supply decisions of workers using data generated in an online labor market. The labor task requires workers to transcribe information from gasoline receipts, and employer's race and sex are signaled via a photograph of a hand holding receipts. Our empirical analysis reveals several findings. First, workers are more likely to correctly identify the race and sex of white and male employers compared to female and black employers. Second, we find mixed evidence of discrimination on the extensive margin. On the one hand, those who correctly identify the race (sex) of the employer are more likely to work for black (female) employers compared to white (female) employers. On the other hand, those who incorrectly identify race (sex) are less likely to work for black (female) employers compared to white (male) employers. Third, white workers transcribe fewer pictures and do so less accurately for black employers relative to white employers. Fourth, both male and female workers transcribe more pictures and do so more accurately for female employers relative to male employers. Finally, results from a survey of mTurkers suggest that the gaps we identify are not driven by statistical discrimination.

JEL Classification: C93, J7, J22

Keywords: labor market; gender discrimination; racial discrimination; mTurk

*Philipp Doerrenberg, University of Mannheim Business School, ZEW, MaTax, IZA, and CESifo. Email: doerrenberg@uni-mannheim.de.

[†]**Corresponding Author:** Denvil Duncan, O'Neill School of Public and Environmental Affairs, Indiana University, SPEA 375F, 1315 East 10th Street, Bloomington, Indiana 47403, U.S.A. (email: duncande@indiana.edu, phone: 812-855-7493).

[‡]Danyang Li, Department of Economics, Hofstra University, Barnard Hall 200E, Hempstead, NY 11549 USA. (email: Danyang.Li@hofstra.edu, phone: 516-463-5599).

1 Introduction

The existing literature on discrimination in labor markets tends to focus on employer \rightarrow employee and employee \rightarrow employee relationships. The employer \rightarrow employee studies cover cases where employers make job-related decisions (e.g., job, promotion, and salary increases) using workers' characteristics such as sex and race rather than workers' productivity. Numerous audit and experimental studies have confirmed race and gender in this branch of the literature (Bertrand and Duflo 2017; Neumark 2018). The employee \rightarrow employee studies focus on cases where employees discriminate against each other. This could arise among employees of similar rank (e.g., Hedegaard and Tyran (2018)) or employees of different ranks (e.g., Glover et al. (2017), Abel (2019)). While both branches have highlighted the desperate treatment of minority groups in the labor market, it is important to recognize that another dimension of discrimination in labor markets has not received as much attention. In particular, we are not aware of any study that documents discrimination by employees against employers. Studies that focus on discrimination between employees of different ranks move us in this direction, but do not fully capture the various margins of discrimination against an employer.

We contribute to the literature by asking the following question: do workers discriminate against employers on the basis of the employer's race or sex? Data for the analysis are generated in a field experiment that was conducted on the Mechanical Turk (mTurk) labor-market platform. A total of 2174 subjects participate a human intelligence task (HIT), which asks subjects to complete a survey. Upon completion of the survey, subjects are randomly assigned to one of five groups and offered the opportunity to complete a bonus task where they are paid a piece rate to transcribe information from gasoline receipts. The instructions to the bonus task features a photograph of a hand holding a gasoline receipt and we vary the presence and characteristics of the hand in the photograph across groups.¹ Subjects are asked if they wish to complete the bonus task. Subjects who respond yes are allowed to transcribe up-to 40 receipts before completing a post-experimental survey. The survey includes two questions, which allow us to define the saliency of the treatment. A subject is classified as treatment-salient if her perceived treatment matches her actual treatment.

With these data, we estimate the impact of the employer's race and sex on three margins of workers' labor supply responses: exit probability², number of receipts transcribed, and transcription quality. In each case, the treatment effect is defined as mean

¹The photograph used in the control group does not feature a hand while the photograph used in the treatment groups features either a black or white hand with or without nail polish.

²To be more specific, we examine the impact of treatment on the likelihood that a worker exits the labor task. Also, notice that we are only able to comment on the effect of treatment on the worker's labor supply to this specific employer because we do not know what workers do when they exit our labor task.

outcome for minority employer (black or female) minus mean outcome for majority employer (white or male).

Our analysis reveals an interesting pattern of discrimination. First, we find no evidence of discrimination for either race or sex on the exit probability using the full sample; i.e., estimated average intent-to-treat effect is zero. When we split the sample by salience of treatment, we find that sex-salient workers were 13.5 percentage points ($p - value < 0.001$) more likely to work for female employers relative to male employers while sex-nonsalient workers were 9 ($p - value = 0.007$) percentage points less likely to work for females relative to males. We find a similar pattern of results when we estimate the race gap; intent-to-treat effect is zero with a race-gap of 11 percentage points ($p - value = 0.001$) for race-salient workers and -4 percentage points ($p - value = 0.31$) for race-nonsalient workers.

There does not appear to be much within-group bias on the race margin as both white and non-white race-salient workers were more likely to work for the black employer relative to the white employer.³ The evidence for within-group bias on the sex margin is mixed. On the one hand, both male and female sex-salient workers were more likely to work for female employers; sex-gap is 13 percentage points for both groups ($p - values < 0.007$). On the other hand, male sex-nonsalient workers were more likely to work for male employers; sex-gap is -13 percentage points ($p - value = 0.006$). We do not find any meaningful race-gap on the intensive margin (number of transcribed receipts).⁴ However, there is a modest sex-gap on the intensive margin that is driven entirely by male workers who transcribe 2.5 to 3 additional receipts for female employers relative to male employers. Female workers transcribe at the same rate for both male and female employers.

Finally, we find evidence of race and sex gaps in the effort exerted by workers as measured by the accuracy of the transcriptions. Workers were more accurate for female employers and less accurate for black employers. The race gap is driven almost exclusively by white workers who were approximately 7 percentage points more accurate for white employers relative to black employers. The estimated sex-gap in accuracy is about 5 percentage points among male and female workers, which suggests within-group bias among females and out-group bias among males.

Overall, our results suggest that workers consider potential employers' race and sex when making labor supply decisions. But, do workers discriminate against employ-

³The non-white sample is small so we are careful not to place too much emphasis on the within-group effects among the non-white sample.

⁴We find suggestive evidence of within-group racial bias on the intensive margin. In particular, white workers transcribe approximately 2 fewer receipts for black employers relative to white employers, while non-white workers transcribe about 2 more pictures for black employers relative to white employers. However, the number of non-white workers in the sample used in the intensive margin analysis is small; approximately 174. Therefore, we are careful not to emphasize the within-group results for the non-white sample of workers.

ers because of taste or statistics? We argue that the primary channel through which statistical discrimination could arise in our setting is through workers’ beliefs about the likelihood that employers of a certain type would honor the labor contract.⁵ For example, a worker might be less likely to work for black or female employers if they believe that these employers are less likely to approve and pay the bonus.

We check for the source of discrimination by surveying a sample of mTurkers. The primary goal of the survey is to collect information about workers’ interactions with employers including workers’ perceptions about employers of different types. Results from this survey suggest that the sex and race gaps we estimate are not driven by statistical discrimination. For example, we find that mTurkers believe male and female employers are equally likely to pay a bonus that was specified in a HIT. The evidence against statistical discrimination is even stronger in the case of the race gap. Here we find that workers were more likely to work for a black employer despite the fact that mTurkers generally believed that black employers are less likely to pay a bonus.

Our paper contributes to the literature on labor-market discrimination by exploring the extent to which employees discriminate against employers. There have been numerous empirical studies of discrimination in labor markets (see [Neumark \(2018\)](#), [Bertrand and Duflo \(2017\)](#), [Riach and Rich \(2002\)](#) for recent reviews). While some studies focus on employees discriminating against their bosses or other employees ([Glover et al. 2017](#); [Hedegaard and Tyran 2018](#); [Abel 2019](#); [Benson et al. 2019](#)), the existing literature mostly studies discrimination by employers toward workers.

Traditional labor markets where workers seek employment might make discrimination against workers an interesting area to investigate. Workers in these cases are interested in working for a firm and the firm in most cases is a neutral entity with respect to race or gender.⁶ Therefore, the worker might not be particularly interested in discriminating against the employer.⁷ However, there are many labor-market arrangements where workers are more accurately described as short-term contractors. Workers in these settings work directly for a specific person (the employer) and are therefore able to consider the employer’s characteristics when deciding whether to take on a given task. These types of employment arrangements have increased significantly in the last decade with the rise of the ‘gig’ economy. Therefore, it is important to understand the extent of

⁵Although requesters are required to hold funds in an Amazon account prior to publishing a HIT, a worker is only paid after the requester approves that worker’s work. Additionally, approval of a task does not guarantee that a bonus will be paid because the bonus payments are not held in escrow. In other words, the requester must first approve the work and then process each worker’s bonus separately.

⁶That is to say, the worker hardly considers the owner of the firm when making a decision to apply for employment. Furthermore, in many cases the owner of the firm is not a single person, but instead many shareholders of varying types including other firms.

⁷Of course a worker might still express a preference for working for a supervisor of the particular race and sex ([Glover, Pallais, and Pariente 2017](#))

discrimination in these settings. Our results suggest that discrimination in these contexts might not be as straight forward as in the previous literature.

We also add to an extensive literature that examines discrimination in platform/online markets. These include markets such as housing rental on AirBnB (Edelman et al. 2017), ride-share (Ge et al. 2016; Cook et al. 2018), and consumer markets (Pope and Sydnor 2011; Nunley et al. 2011; Doleac and Stein 2013; Zussman 2013; Ayres et al. 2015). Our experimental context is an online labor market where workers complete micro tasks for pay on a contractual basis. Unlike many of the other market platforms that have been studied so far, race and gender are not particularly salient in our setting. This is a particularly interesting case to study because it allows us to comment on the likely effects of increasing the saliency of employer characteristics. Interestingly, we find that minority groups might benefit on some margins but not others. This is unlike the existing literature, which finds almost unanimous evidence that minority groups face discrimination when race and sex are salient.

One implication of our findings is that crowd-source labor markets that are designed with strong contract-based arrangements like mTurk can minimize their requester’s exposure to discrimination on the basis of race and sex by reducing the saliency of these characteristics. While mTurk maintains a strong sense of anonymity, this is not true of all similar labor markets. For example, some of these labor markets require both requesters and workers to establish user-profiles with names and pictures.

The remainder of the study proceeds as follows. We discuss the experimental design in section 2. This is followed by a description of the data in section 3, empirical strategy and results in section 4, and discussion in section 5. We conclude in section 6.

2 Design and implementation

Our objective is to determine whether workers consider employers’ race or sex when making labor supply decisions. We isolate the effect of employer’s race and sex on workers’ labor supply via a field-experiment on a crowd-sourcing labor platform. The remainder of this section provides a detailed description of the experimental design.

2.1 Design

Recruitment. We recruit subjects from Amazon’s Mechanical Turk (mTurk) using a HIT that invites subjects to complete a road mileage user-fee survey on Qualtrics for a flat fee of \$0.65. Upon completion of the survey, subjects complete a brief demographic-questionnaire and are then randomly assigned to one of five treatment groups that differ only in the race/sex signal of the employer that we send to subjects. Figure 1 provides an illustrative diagram of the flow of the experiment.

Treatment. Once subjects are assigned to treatment groups, we thank them for completing the mileage user-fee survey, and inform them that there is an opportunity to earn additional income by completing a transcription task for the requester. The task and its description to subjects are identical across treatment groups; subjects have to transcribe gas station name, date of purchase, gallons of gasoline purchased, price per gallon, and total sale value from gasoline receipts hoarded by one of the authors. We also tell them the approximate time it will take to transcribe the information from one receipt (30 seconds) and the wage per receipt (\$0.06). See Figure 2 for a screen shot of the details shown to subjects at the time they receive treatment.

Subjects are shown pictures of the gasoline receipts (see Figure 3). As indicated above, the receipts are identical across treatment groups except for the signal of race and sex. Following Doleac and Stein (2013), we signal race and sex by showing subjects a picture of a hand holding a receipt. We start with a stock of gasoline receipts that one of the authors collected over a four year period. From this stock of receipts we selected approximately 100 receipts that were in good condition; all of the information we wanted subjects to transcribe was visible. We then selected four hand models; one black and one white female, and one black and one white male. In order to make race and sex salient, we selected black hand models with dark skin, and asked the females to wear nail polish. We then conducted a photo-shoot where we took a picture of each person holding each of the receipts; approximately 100 pictures per person. The pictures included only the receipt and the model’s hand. Finally, we selected the most clear pictures for each hand model. From this list, we identified the set of clear receipts that were common across models. This left us with 40 receipts, which were used in the experiment.

Therefore, subjects in the black female (BF) group see receipts held by a hand with black skin and nail polish. Subjects in the black male (BM) group see receipts held by a hand with black skin (no nail polish). Subjects in the white female (WF) group see receipts held by a hand with white skin and nail polish. Subjects in the white male (WM) group see receipts held by a hand with white skin (no nail polish). Finally, subjects in the control group see receipts that do not include a hand. Again, the receipts available for transcription are identical across treatments and the receipts are presented in the same order across groups.

We ask subjects if they would like to transcribe the receipts after exposure to the treatment pictures (see Figure ??). Because the transcription was not included in the recruitment HIT, we make it clear to the subjects that the transcription task is optional and that there is no penalty for opting out. Subjects who respond yes, transcribe receipts sequentially and are allowed to exit the task after each receipt (see Figure 5 for details).

Salience of requester’s race and sex There are three sources of saliency to consider. First, it is important that subjects do not select the mTurk HIT based on their perception of the requester’s sex or race as signaled by the requester’s name because we only have data for workers who accept the HIT. We select a requester-name ‘Alex Wright’, which is mostly neutral with respect to race and sex. This approach should minimize the likelihood that subjects select the HIT based on the requesters race or sex. We can confirm that the demographic characteristics of our sample is very comparable to that of other samples that the authors and other researchers have recruited from mTurk in the past.⁸

Second, we want to make sure subjects receive the signal we intended to send via the hands in the pictures. The salience of the race and sex signals were tested in a pilot-experiment. We found that the majority of subjects correctly identified the race and sex of the hands.⁹ The treatment is presented in a way that should maximize the salience of the hands. After subjects complete the mileage survey and click submit, they are taken to a new page that has the picture of the receipt at the top of the page. Depending on the size of the subject’s screen, the picture with the receipt and the hand will be the only thing the subject sees before scrolling down the page. We follow-up this design feature with a survey at the end of the experiment to capture subjects perception of their treatment status. Subjects are asked about the race and sex of the person in the picture they saw. We also ask subjects about the United States president in order to check if subjects were paying attention. Responses to this post-treatment survey are used to determine whether subjects correctly perceived their treatment status.

Finally, we want to make sure subjects make the connection between the hand in the picture and the requester. We attempt to increase the connection between the hand and the requester by writing the mTurk HIT and the treatment in first-person singular ‘I’. For example, the mTurk HIT includes language like “I would like your opinion about the move toward the mileage tax.” Similarly, the instructions subjects see when they receive treatment is written with the intent of connecting the hand to the person making the job request; see 2. For example, we tell subjects “*I want to know how much I would pay in mileage tax compared to what I now pay for gasoline tax*”, “*I would like you to transcribe information from **my** gasoline receipts*”, and “*I have included a sample of*

⁸On average, our sample is 78% white, 56% with B.Sc. or higher, 37 years old, and 48% female (see Table 1). This is comparable to the US sample in Bohren et al. (2019) and the samples in Duncan and Li (2018) and Kuziemko et al. (2015a).

⁹The results of a pilot study that specifically checked whether subjects could identify the race and sex of the hands found that 83% of subjects correctly identified the race and sex of the black-female hand; 62% and 75% correctly identified the race and sex of the black-male hand, respectively; and 90% of subjects correctly identified the race and sex of the white-female hand. The race of white-male hand used in the pilot was correctly identified 86% of the time, but the sex was only correctly identified 25% of the time. We changed the white-male hand model for the actual experiment, but did not run another pilot. However, the race and sex of the white-male hand used in the actual experiment was correctly identified by 79% and 69% of the subjects, respectively.

one of *my* receipts above”.

2.2 Implementation

The experiment was conducted on Qualtrics using subjects recruited from mTurk. We first create a human intelligence task (HIT) that is advertised on mTurk. The HIT includes a description of a survey and compensation. We deliberately exclude any mention of the transcription task in the HIT. Introducing the transcription task directly on mTurk means that we would not be able to say anything about the demographics of the subjects who view, but do not accept the treatment. Additionally, we would only be able to track those who accepted the HIT, which means that we wouldn’t be able to say much about acceptance rates, which is one of our primary outcome of interest. Instead, we recruit a large sample of subjects to complete a survey and then introduce the treatment. In this way we are able to collect demographic variables on all of our subjects even if they subsequently refuse to transcribe the receipts.

Subjects are told to accept the HIT and click on the weblink if they are interested in completing the survey. Subjects who click on the link are taken to our Qualtrics site where they complete the survey before being assigned to a treatment group to transcribe images. We selected the mileage user-fee survey and gasoline receipt transcription task because it allowed us to present the whole experiment as one event being implemented by a private citizen who is concerned about her state potentially adopting a road mileage user-fee. This reduces the likelihood that subjects view the HIT as part of an academic study thus preserving the reliability of their decisions and responses. Transcribing text from a scanned or photographed receipt is a common type of task on mTurk. This further reduces the chances that subjects realize they are participating in an experiment.

We chose to run the experiment on mTurk for several reasons. First, mTurk is one of the largest online labor markets where job offers are posted and workers choose jobs for payment. According to Amazon, there are over 500,000 workers from 190 countries in the mTurk labor market: <https://requester.mturk.com/tour>. Therefore, mTurk has a special place in the digitally-mediated labor markets that have come on the scene in the last decade. Second, experimenter effects are avoided because subjects do not know that they participate in an experiment (Paolacci, Chandler, and Ipeirotis 2010; Horton, Rand, and Zeckhauser 2011; Buhrmester, Kwang, and Gosling 2011; Mason and Suri 2011). Importantly for us, we are able to identify the effect of race and sex in a naturally occurring labor market. In general, experiments on Amazon’s Mechanical Turk therefore combine internal and external validity since it is a “real” labor market with actual workers where randomized trials can be conducted (Horton, Rand, and Zeckhauser 2011).¹⁰

¹⁰Kuziemko et al. (2015b) and DellaVigna and Pope (2018) are recent examples of economics papers using Amazon’s Mechanical Turk.

Payment The experiment ends for each subject when she decides to stop or when she transcribes 40 pictures, whichever comes first. In either case, each subject is instructed to copy her personal ID number and paste it in the entry box on the mTurk website. This process is necessary for us to match subjects to their mTurk worker ID and thus process their payments. Subjects receive a participation reward of \$0.65, which is paid as long as a subject accepts the HIT and completes the survey. Additionally, subjects are paid a piece rate of \$0.06 for each transcribed receipts. Given the payment restrictions imposed by the mTurk platform, we frame the piece rate as a bonus in all communications to the subjects. Overall, we paid a total of \$2419 for 2500 subjects who took an average of 7.8 minutes to complete the study; this translates to an hourly effective wage of approximately \$7.4.¹¹

3 Data Summary

3.1 Data Cleaning

We fielded the experiment in two waves collecting 1250 responses each time for a total of 2500 subjects; approximately 500 observations per treatment group.¹² We cleaned the data in the following ways before performing our empirical analysis. First, we calculate the total time taken to complete the experiment in minutes and trimmed the top 5% and bottom 5% of the sample. This removed subjects who took fewer than 2 minutes or more than 38 minutes; 267 subjects uniformly distributed across our treatment groups. Second, we drop 66 subjects who stated that the president of the US is Michael Jordon since this is an indication that subjects were simply clicking through the study. Finally, we drop 68 cases where subjects had the same ipaddress because this might be an indication that the same subject is taking the experiment multiple times or it could be that turkers from other countries are taking the experiment when they should not. These adjustments leaves us with 2174 total observations; a bit over 430 subjects per treatment.

3.2 Demographic Characteristics

Because we are interested in race and sex discrimination and the groups are very similar to each other on observables (see appendix Table 7), we combine the groups to form race and sex groups. Summary statistics for these race and sex combinations are presented in Table 1. Overall, our sample is typical of other mTurk samples; average age of 37, 78% white, 48% female, 51% urban, and highly educated with approximately two-thirds of

¹¹This payment dos not include the administrative fee paid to Amazon.

¹²As indicated above, the HIT included two parts: a mileage user-fee survey and a transcription task. The current paper analyses the data from the transcription task. The mileage user-fee data are used to write a separate paper on public opinion of mileage userfees.

subjects having at least a two-year college degree.

For ease of comparing demographic characteristics across groups, we take the difference in means between treatment and control groups for each demographic variable and present these results in Table 2 along with p-values from a ranksum test of the null hypothesis that the means are the same. The majority of the differences are statistically indistinguishable from zero. Notable exceptions are race, sex, and education where we observe small differences between the treatment and control groups in some cases. More importantly, except for education, there is no statistically significant difference between the female and male treatments. We find that, relative to the male treatment, the female treatment has 4 percentage points fewer subjects with a B.Sc. degree and 4 percentage points more subjects with a Graduate degree; $p - value = 0.089$ & 0.011 , respectively. Similarly, subjects' race is the only statistically significant difference between the black and white treatment groups; 4 percentage points more white subjects in the black treatment relative to the white treatment. We control for these variables in the empirical analysis and find that they do not change our results.

3.3 Salience of Treatment

One of the advantages of our experimental design is that we collect data on subjects' perception of their treatment status with two post-experiment questions. The two questions are:

1. What is the race of the person holding the receipt in the picture?
2. What is the sex of the person holding the receipt in the picture?

Possible responses are black/female, white/male, I don't know, the picture did not include a person. We randomized the order of the questions and the answers to control of any order effects. We use responses to these questions to create a measure of treatment-salience. The salience measure indicates if subjects' perceived treatment is equal to their actual treatment, where perceived treatment is based on the subjects' responses to the questions above. For example, a subject is labeled race-salient if her response to question 1 is correct, and sex-salient if her response to question 2 is correct.

These measure of salience is used to infer whether or not a subject was actually "treated". Of course, this is not a perfect measure of treatment since it is self-reported, which opens the possibility of misreporting. However, it is instructive to see whether the treatment effects are conditional on treatment salience. Figure 6 shows the share of subjects whose perceived treatment matches their actual treatment, for race and sex. We find that just under 40% of subjects in the black treatment correctly perceived their treatment compared to over 80% of subjects in the control and white treatment groups.

Similarly, subjects in the male treatment were more likely to correctly perceive their treatment status than subjects in the female treatment; 65% for males versus 47% for females.¹³ The summary statistics in Table 3 show that the demographic profile of subjects is mostly similar across race and sex salience. Subjects for whom race was salient tended to be modestly younger and from urban areas, while sex-salient subjects tended to be modestly younger.

4 Empirical Strategy and Results

This section describes our empirical strategy and results. We present extensive-margin results followed by results for the intensive margin and accuracy. In each case we present intent-to-treat and treatment-on-treated effects.

4.1 Empirical Strategy

We estimate equation 1 to determine if subjects consider requesters' race and sex when making their labor supply decision in the transcription task.

$$y_i = \alpha + \beta Treatment_i + \delta X_i + \epsilon_i, \quad (1)$$

where y_i is one of three outcome variables; transcribe, number of receipts transcribed, and accuracy. Transcribe is an indicator variable that takes a value of 1 if the subject accepted the transcription task and zero otherwise. Accuracy is measured by the share of accurate entries (more on this below). *Treatment* indicates subjects' treatment assignment. For example, in the race specifications, *Treatment* is equal to 1 if the subject was assigned to the black-hand treatment and zero if the subject was assigned to the white-hand treatment.¹⁴ Therefore, β is the estimated race or sex gap depending on the specification; positive values indicate that discrimination benefits the minority group (black or female employers). X is a vector of covariates including age, sex, race, education and urban, and ϵ_i is a random error term. We estimate equation 1 on the full sample of subjects to obtain intent-to-treat estimates of discrimination, and on the various salience sub-samples to obtain treatment-on-treated effects.

¹³Results presented in Figures 25 to 30 show subjects' responses across the possible responses on the post-experiment race and sex questions. We find that just under 40% of subjects in the black treatment correctly perceived their treatment, while 30% reported being in the white treatment, 7% reported being in the control group, and 24% did know the race of the hand in the picture. On the other hand, over 80% of subjects in the control and white treatment groups correctly perceived their treatment status with the remaining subjects mostly saying they don't know the race of the hand. The findings are somewhat similar when we look at the salience of the sex treatments. Subjects in the women treatments are more likely to misperceive their true treatment status.

¹⁴We exclude the control group in these specifications. However, the result we obtain is the same as if we estimated $transcribed = \alpha + \beta_b black + \beta_w white + \epsilon$ and then calculate $\beta = \beta_b - \beta_w$.

4.2 Results

4.2.1 Extensive Margin

Intent-to-Treat Figure 7 reports the acceptance rate across race and sex groups. The figure shows that the mean acceptance rate was approximately 36% across treatment groups.¹⁵ Importantly, there does not appear to be much difference in subjects' willingness to transcribe receipts across employer characteristics.

We estimate equation 1 to check whether workers decision to transcribe receipts was influenced by the employers race or sex. The estimated β 's for each specification (race or sex) and samples are presented in Figure 8 for race and sex. The first panel of results, which are for the full sample, show no evidence of bias based on race or sex of the employer. The estimated coefficients are practically zero for both race and sex, which indicates that, on average, subjects were equally likely to work for black/female employers as white/male employers.

Treatment-on-Treated The remaining results in Figure 8 show that the null effect described above hides an interesting pattern in the data. In particular, we find large and statistically significant effects when we split the sample based on the salience of the treatment. Subjects for whom the race treatment was salient were 11 percentage points more likely to work for a black employer than a white employer. Similarly, sex-salient subjects were 14 percentage points more likely to work for a female employer than a male employer.

Our null result in the full sample combined with very large treatment effects among the 'treated' subjects suggests very large negative treatment effects in the non-salient sub-samples. This is precisely what we find. As shown in Figure 8, subjects who "mis-perceived" their treatment were less likely to work for black or female employers. The estimated gap among the non-salient sub-sample is -9 percentage points (p-value=0.007) for sex and -4.3 percentage points (p-value=0.312) for race. This finding is puzzling since we would have expected to find a null result among subjects who paid no attention to the race or sex of the employer. We explore this finding further in Section 5.

Group Dynamics: race We explore with-in group dynamics by cutting the samples by the race of workers when estimating the race gap and by the sex of workers when estimating the sex-gap. We classify workers as either white or non-white based on their responses to the the post-experiment survey. The non-white group is a fairly small share (only 22%) of the total sample. Therefore, we are careful when interpreting the race

¹⁵Figure 34 shows that including a hand did not affect the extensive margin decision to transcribe receipts. Receipts with a hand had an acceptance rate of approximately 37% compared to 35% for the control group. See Figure 35 for detailed results across treatment groups.

estimates, especially for the subsamples on race salience. Even so, we find a similar pattern of results when we split the sample by the race of the workers. Figure 9 shows that while there is no evidence of a racial gap in the full sample, both white and non-white workers for whom the employer’s race was salient exhibited a strong preference for working for the black employer. This result also shows that the preference for working for a black employer is stronger among non-white workers. Interestingly, we find no statistically significant evidence of a racial gap among workers for whom the employer’s race was not salient. However, there is an economically substantive gap among non-white workers for whom race was not salient; the gap among white workers is both economically small and statistically indistinguishable from zero.

Group Dynamics: sex Figure 10 reports similar within-group results for the sex-gap. In particular, we find no evidence of a gap in the full sample for neither female nor male workers, but a large positive and statistically significant gap among both male and female workers for whom the employer’s sex was salient. Interestingly, we find a strong negative sex-gap among male workers for whom the employers sex was not salient. The gap is similarly negative among female workers in this subsample, but the estimate is much smaller and cannot be distinguished from zero.

Overall, our results suggest that women generally prefer working for women. While some men prefer working for women, other men prefer working for men. The group-dynamics for race are also interesting. There appears to be out-group bias among white workers who are more likely to work for black employers relative to white employers. The evidence is more mixed for non-white workers. Nonwhite workers for whom race is salient prefer working for black employers while non-white workers for whom race was not salient prefer to work for white employers. Again, the small number of non-white subjects makes us cautious in interpreting the race group-dynamics.

4.2.2 Intensive Margin

Approximately 36% (or 753) of subjects transcribed at least 1 receipt, and subjects transcribed an average of 7.6 receipts. However, the distribution is highly skewed; the median number of transcribed receipts is 3, 75% of subjects transcribed fewer than 9 receipts, 90% transcribed fewer than 21 and only 37 subjects transcribed all 40 receipts. Figure 11 shows how the mean number of transcribed receipts varies across treatment groups. Subjects in the control group transcribed 10 receipts on average. While subjects in the treatment groups transcribed fewer receipts than those in the control group, the reduction appears to be uniform across the race treatments and slightly larger for male employers compared to female employers.

We estimate intensive margin effects using equation 1 with ‘number of transcribed

receipts' as the outcome variable. The model is estimated separately for the full sample of transcribers (those who transcribed at least one receipt) and those transcribers for whom treatment was salient. The results presented in Figure 12 show that workers transcribed more pictures for female employers relative to male employers and fewer receipts for black employers relative to white employers. However, the differences are small and statistically indistinguishable from zero. The results for the non-salient group is in line with expectations; the estimated gap is practically zero. A more detailed analysis reveals that there is no race gap among female employers; workers complete the same number of transcriptions for both black and white female employers. There is a gender gap for both races; subjects completed more transcriptions for female employers than male employers regardless of employers race.

There is suggestive evidence of within-group bias for the race-gap and out-group bias for the sex-gap on the intensive margin. In particular, we find that non-white workers complete more transcriptions for black employers than white employers, while white workers complete more pictures for white employers than black employers (see Figure 13). On the sex margin, Figure 14 shows strong evidence that male workers complete more transcriptions for female employers than male employers. However, female workers expressed no preference for the gender of the employer when deciding how many receipts to transcribe.

4.3 Accuracy

This section describes the accuracy of the transcriptions. We define accuracy as the share of correct entries across all transcribed receipts for each subject and test whether accuracy differs across employer's race and sex.

Definition of Accuracy Each receipt had seven items for subjects to transcribe. Let n_i be the number of receipts transcribed by subject i . Then the total number of items transcribed by subject i is $T_i = 7 * n_i$. If we define c_i as the number of correct items for subject i , then the accuracy rate for subject i is $a_i = c_i/T_i$. The first step in creating this variable is to transcribe the receipts ourselves. Second, we compare our transcription of each item to the corresponding transcription for each subject and adopt two separate rules to identify correct entries. The first rule is strong in the sense that an entry is correct if it is an exact match to the corresponding entry on the receipt. This decision rule does not allow for rounding of dollar figures. However, the receipts report dollar figures to three decimal places and we do observe that some subjects round these entries. Further, there is a lot of variation across subjects in the rounding rule; some round to two decimals place, others to one decimal place and so on. Because we did not include any instructions about rounding in the experiment, we adopt a weaker definition of accuracy,

where an entry is labeled accurate if it matches the corresponding entry on the receipt or any of its possible rounded representations. So, if the receipt lists price at \$2.476, then \$2.476, \$2.48, \$2.5, and \$2 would all be coded as accurate under weak accuracy, while only \$2.476 would be coded as accurate under strong accuracy. Following this procedure, we calculate $a_i = c_i/T_i$ for each subject.

Treatment effect Figure 15 summarizes the accuracy of transcription across race groups. We find that subjects in the control group got about 87% of their entries correct. The corresponding rate for the treatment groups is 76% and 80% for black and white groups, respectively, and 80% and 75% for female and male employers, respectively. The estimated treatment effects presented in Figure 16 show that these differences are statistically different from zero for both race and sex and across the subsamples. In particular, we find that, conditional on transcribing receipts, workers are more accurate for female employers and less accurate for black employers.

Figure 17 shows evidence of within-group bias among white workers who were more accurate when working for white employers compared to black employers. There is no evidence that non-white subjects' accuracy level varied with the race of the employer. The results in Figure 17 show that both male and female workers were more accurate when transcribing for female employers compared to male employers.

5 Discussion

Our results suggest that mturk workers consider a requesters race and sex when deciding when making labor supply decisions. We also find some evidence that there are two type of subjects; those who are biased toward minority groups and those who are biased against minority groups. This section of the paper explores these findings. We first discuss the case for the heterogenous preference toward minority groups. Next we explore the extent to which the sex and race gaps reflect an underlying preference for certain types of employers versus statistical discrimination.

5.1 Two types of subjects

We find that salient and non-salient subjects respond differently to the treatment suggesting that our salience variable identifies two types of subjects. The fact that we find evidence of discrimination among subjects in the non-salient group is particularly interesting and open to multiple explanations. We explore this pattern in the data and discuss possible explanations in this subsection.

Figure 3 shows the race and sex signals that we sent to subjects (actual treatment) and Figure 6 shows subjects' perception of those signals (perceived treatment).

Approximately 80% of subjects in white treatments correctly perceived treatment status compared to only 37% in the black treatments. Interestingly, 31% of subjects in the black treatments reported that the employer is white, while 25% stated they did not know the race of the employer. Only 0.71% of subjects in the white treatments stated that the employer was black and 11.6% responded that they did not know the race of the employer. A similar pattern is observed for sex; 65% of subjects in the male treatments correctly identified the sex of the employer compared to only 43% in the female treatments. Additionally, 17% of subjects in the female treatments stated that the hand belonged to a male while only 1.8% of subjects in the male treatments said the hand belonged to a woman.

Considering that the signals in Figure 3 are clear, then there are at least three possible explanations for the observed difference in subjects' perception of the employer's race and sex. First, it could be that subjects paid attention to the pictures and correctly identified the race and sex, but ex-post misreported the race and sex in an effort to conceal their biases. Second, it could be that subjects received the race/sex signal and responded subconsciously in a manner that favors majority groups even if they did not pay careful attention to the treatment. This type of channel is referred to as implicit bias in the literature (Bertrand, Chugh, and Mullainathan 2005). Finally, subjects could have ignored the treatment signal and thus failed to respond to the signal, and then guess at treatment in the post-experiment survey.

The pattern of results presented in Figure 8 is consistent with the first and second explanations. In particular, if subjects were inattentive to the treatment and simply guessed a response to the race and sex questions, then there should be no treatment effect among inattentive subjects. This is especially true since attentive and inattentive samples have very similar demographic profiles. Our results are not consistent with this predicted null-treatment effect. Although not statistically different from zero, subjects for whom race was not salient were approximately 5 percentage points more likely to transcribe for white than black employers. The effect is even more dramatic if we exclude subjects who responded 'I don't know'; in this case, non-salient subjects are 15 percentage points more likely to work for white employers than black employers.¹⁶

We also find that inattentive subjects are almost 10 percentage points more likely to transcribe for male employers than female employers. Unlike the race treatments, we also find a large difference among subjects who responded 'I don't know'; 25% ($N = 270$) transcribed in the female treatment compared to 34% ($N = 214$) in the male treatment.

¹⁶Subjects who responded 'I don't know' transcribe at the same rate for both white and black employers; approximately 27% ($N = 213$ in the black treatment and $N = 98$ in the white treatment). Subjects in the black treatment who responded that the employer was white or that there was no hand in the picture transcribe at rates of 26% ($N=262$) and 32% ($N = 56$) for white and 'no hand', respectively. Subjects in the white treatment who responded that the employer was black or that there was no hand in the picture transcribe at rates of 33% ($N = 6$) and 43% ($N = 54$) for black and 'no hand', respectively.

Furthermore, of the 842 subjects assigned to the female treatments, 147 reported that the hand belonged to a male employer and only 14% of these subjects agreed to transcribe images. Only 15 of the 849 subjects in the male treatment reported that the employer is a female.

Overall, these results are very suggestive of a split sample. Some subjects are biased toward minority groups and are not afraid to expose their bias. Other subjects are biased against minority groups either consciously or sub-consciously.

5.2 Tastes or Statistical

In this section we present an exposition of evidence that is supportive of both implicit bias and taste-based explanations of the race and sex gaps. Because mTurkers complete HITs for pay, the most likely source of statistical discrimination in our context is a worker’s expectation of being paid by the employer. Taste-based discrimination, on the other hand, simply requires that the worker is able to identify the employer’s race or sex. However, the mturk labor market is largely anonymous in that workers never meet employers. Although most HITs are accompanied by the employer’s name, which may signal the employer’s race and sex, it is not clear if workers generally pay attention to the requester’s name when selecting HITs. Therefore, we designed a survey of mTurkers to obtain information on workers’ past experiences with employers. The goal is to determine if mTurkers’ perception of likely non-payment is correlated with the race or sex of the employer. We are also interested in identifying the extent to which workers contact employers, pay attention to employer’s name, and whether workers consider employer’s characteristics when selecting a HIT.

5.2.1 Survey Design

The survey has four sections. First, we ask subjects to report their age, sex, race, and education. Second, we ask about their usage of mTurk; year they joined mTurk, whether mTurk is their primary job, and number of HITs completed per month. The third section asks about their experiences as an mTurker. Here we ask about frequency with which work is accepted by requester, experience with requesters’ refusal to pay for work completed, communication with requester, attentiveness to requesters’ name, and whether knowledge of a requester’s characteristics would affect likelihood of accepting a HIT. We also ask subjects about the likelihood of a requester paying a bonus for a completed task.

For the final section, we randomly assign subjects to one of five groups that correspond to the five treatment groups of the original experiment. The treatment is presented as a hypothetical scenario. Specifically, the subjects saw the following text: *“In the next set of questions, I am going to ask you about your perception of what Turkers like you are*

likely to do when faced with a transcription task.” Subjects are then asked three questions; what percent of mTurkers would accept the HIT, would you accept the HIT, and how likely is it that the requester would pay the bonus accompanying the HIT. We also ask subjects to identify the race and sex of the person represented by the hand in the picture. The full set of questions is available upon request.

The survey was fielded to 1012 mTurkers who did not participate in the original study. The data are cleaned as follows; we drop all duplicated ipaddresses ($N = 41$) and everyone who identified Michael Jordan as president of the US ($N = 16$). This leaves us with a sample of 955 subjects. Subjects took an average of 4.3 minutes to complete the survey and were paid a flat fee of \$1.

5.2.2 Summary

We find no meaningful nor statistical difference in the observable characteristics between the race treatment groups.¹⁷ There is a statistically significant difference in the two youngest age groups between the control and male treatments, but these differences are small. We also find that the survey sample is similar to the original real-effort sample in age, sex, race, and education. Importantly, the salience of treatment is identical between the experiment and survey samples (see Table 6).

5.2.3 mTurk Experience

Start-Date and Usage Approximately 25% of the subjects report joining mTurk as a worker before 2016, and 13%, 21% and 41% reported joining mTurk in 2016, 2017 and 2018, respectively. Subjects report completing an average of 474 HITs per month in the full sample. Panel A of Figure 36 shows that there is heterogeneity in HITs completed across subjects’ race but not sex; white subjects complete 106 more HITs per month than non-white subjects (Ranksum $p - value = 0.001$), while female subjects complete only 14 more HITs than male subjects (Ranksum $p - value = 0.29$). The mean monthly completed HITs is 488 in the control group, 479 and 462 for black-hand and white-hand treatments, respectively, and 500 and 440 for female-hand and male-hand treatments, respectively. These differences are not statistically distinguishable from zero (see Panel B of 36).

Experience with Requesters Figure 19 presents the results of subjects responses about their experiences as mTurkers. Approximately 25% of subjects report that mTurk is their primary source of employment. The remaining summary information in Figure 19 describes subjects’ experiences with Requesters and are suggestive of both statistical

¹⁷Results for the observable demographic variables are available upon request; age, sex, race, education and time spent on the survey.

discrimination and taste-based discrimination. First, 45% and 53% of subjects report that their HITs are accepted ‘all the time’ and ‘most of the time’, respectively. Approximately 54% report being in a situation where the Requester refused to pay for a completed HIT, and approximately 80% have contacted a requester in the past. Additionally, subjects reported being 67% confident that a requester who offers a bonus task on an external website, such as was the case in our real-effort task, would pay the bonus upon completion of the task. This suggest that there is a significant amount of doubt about payment in the subjects minds as they make their HIT selection decisions. To the extent that concerns about payment is correlated with perceived race or sex, subjects could use this prior experience to form expectations about the honesty of the Requester. Therefore, rather than selecting on the basis of taste, subjects could instead be selecting HITs on the basis of expected payment by the Requester.

On the other hand, 72% report that that they check the names of requester, and 71% of subjects report that they would consider a requester’s characteristics when making HIT selection decisions on mTurk. These responses could support both taste-based as well as statistical discrimination. It could be that subjects use requesters’ name and characteristics in order to identify probabilistically-honest requesters. Alternatively, subjects could possibly use this information to identify groups of requesters they have a deep-seated bias against.

Interestingly, we do not find any meaningful differences in these self-reported experiences across subjects’ race or sex. This suggests that differential experiences across race and sex is not a strong explanation for the group-dynamics we observe in our real-effort experiment.

We explore the possibility of separating taste-based from preference-based discrimination by presenting subjects with the same treatments as in the real-effort experiment and then asking about acceptance and perceived likelihood of being paid. The question about likelihood of being paid allows us to determine whether perception of requesters’ honesty is correlated with requesters’ race, which further allows us to comment on the source of the bias uncovered in our real-effort experiment.

Role of race and sex Subjects’ responses to the three post-treatment questions are summarized in Figure 20. Subjects reported that approximately 46% of other mTurkers would accept the HIT, but only 34% of subjects reported that they themselves would accept the HIT. So, while subjects thought the acceptance rate among other mTurkers is about 10 percentage points higher than what mTurkers actually did, the subjects’ personal acceptance rate is identical to the acceptance rate observed in the experiment.

Importantly for our analysis, subjects reported a 69% likelihood that the requester would pay the bonus. Again, this suggest that subjects have some a mount of uncer-

tainty about being paid at the time they make their HIT-acceptance decisions and this is suggestive of statistical discrimination as a possible explanation for our results. However, Figure 21 shows that this uncertainty about payment is not correlated with the sex of the requester; estimates are both small and statistically indistinguishable from zero across all samples. There is no statistical or economic evidence that the race of the requester is correlated with the uncertainty of being paid in the full sample. However, this null result masks variation within the salience sub-samples. On the one hand, race-salient subjects reported that they thought black requesters were about 5 (p-value 0.13) percentage points less likely to pay than white requesters. On the other hand, subjects for whom race was not salient reported a gap of only 2 (p-value 0.54) percentage points.

The fact that subjects thought both male and female employers are equally likely to pay suggests that the biases observed in our experiment are not driven by statistical discrimination. The results for race are similarly suggestive of an even stronger rejection of statistical discrimination. Notice that subjects in our salient-samples are more likely to work for the black employer despite the general impression among mTurkers that black employers are less likely to pay. Results for the non-salient sample are also suggestive of a taste-based effect. Subjects in this group were less likely to work for the black employer although they do not believe there is any difference in likelihood of being paid.

We extend the analysis to account for the possibility of within-group bias by estimating the treatment effect separately for sex and race of the subjects, and reporting the results in Figures 22 and Figures 23 for race and sex, respectively. While there does not appear to be any statistical evidence for a difference in perceived likelihood of the employer paying among non-white subjects, we find that white subjects in the salient sample tend to believe that black employers are less likely to pay than white employers. These results further support the case for a strong taste-based source of bias; white workers in the salient sample preferred working for black employers despite the perception among their peers that black employers are less likely to pay. Similarly, non-white workers prefer black employers despite no apparent difference in perception of being paid.

We find no evidence of within-group bias in the female treatments. Estimates are small and statistically indistinguishable from zero except in the non-salient sample where estimates are fairly large. Among subjects in the non-salient sample, female subjects report that female employers are about 6 percentage points less likely to pay than male employers, while male workers report that female employers are about 4 percentage points more likely to pay. These results are also inconsistent with statistical discrimination; males in the non-salient sample were less likely to work for female employers despite the fact that their peers are of the view that female employers are more likely to pay than male employers.

Overall, the results from our mTurk user-survey show that both taste-based and

statistical discrimination are possible. However, the results of our real-effort experiment are not consistent with statistical discrimination. Although the results are more consistent with taste rather than statistical discrimination, we cannot rule out implicit bias as a possible explanation. This is particularly relevant in our empirical setting where workers are making decisions very quickly. Because each HIT pays a relatively low wage, workers generally try to complete many HITs in order to make a high income. In this setting, it is likely that a worker’s implicit bias is the primary vehicle through which discrimination arises.

6 Conclusions

We estimate the effect of employers’ race and sex on the willingness of workers to persist on a labor task using data generated on Amazon’s Mechanical Turk. We find strong and persistent evidence that if given the opportunity, workers in contractual-types of jobs will exhibit bias both toward and against female employers. These biases appear to exist among both male and female workers although the bias against female employers is stronger among male workers. We find a similar pattern of results for the race of the employer although the evidence for workers with a bias against black employers is not as strong as that of sex.

Results from a mTurk-user survey suggest that the biases we detect are not driven by statistical discrimination. Although subjects express some uncertainty about the likelihood of being paid for mTurk HITs, this uncertainty is not correlated with the sex of the employer. Furthermore, the correlation with race and likelihood of being paid is not consistent with statistical discrimination. On the one hand, we find that subjects who prefer to work for black employers believe black employers are less likely to pay than white employers. On the other hand, those who prefer to work for white employers perceive no difference in likelihood of paying between black and white employers.

Therefore, to the extent that the likelihood of being paid is the primary channel through which statistical discrimination would manifest itself in our setting, this finding suggest that the biases we estimate are not driven by statistical discrimination.

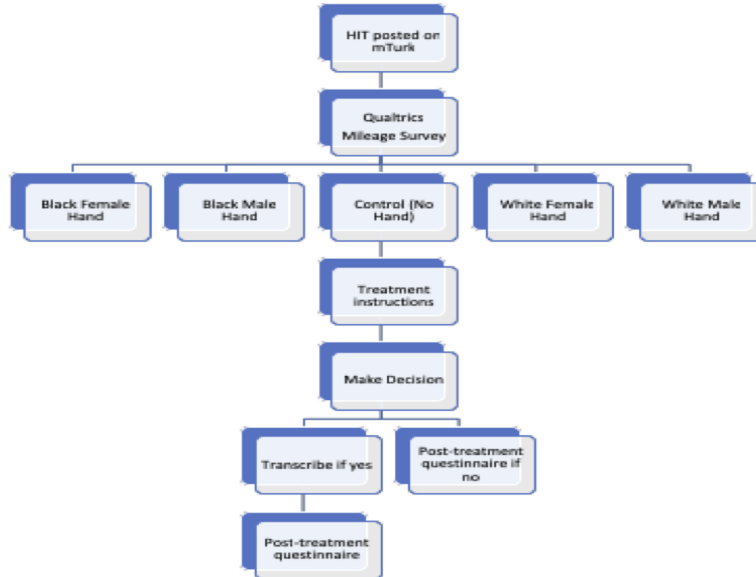
References

- Abel, M. (2019). Do workers discriminate against female bosses?
- Ayres, I., M. Banaji, and C. Jolls (2015). Race effects on ebay. *The RAND Journal of Economics* 46(4), 891–917.
- Benson, A., S. Board, and M. Meyer-ter Vehn (2019). Discrimination in hiring: Evidence from retail sales.
- Bertrand, M., D. Chugh, and S. Mullainathan (2005). Implicit discrimination. *American Economic Review* 95(2), 94–98.
- Bertrand, M. and E. Duflo (2017, jan). Field Experiments on Discrimination. *Handbook of Economic Field Experiments* 1, 309–393.
- Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019, June). Inaccurate statistical discrimination. Working Paper 25935, National Bureau of Economic Research.
- Buhrmester, M., T. Kwang, and S. D. Gosling (2011). Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6(1), 3–5.
- Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (2018). The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers *. Technical report.
- DellaVigna, S. and D. Pope (2018). What Motivates Effort? Evidence and Expert Forecasts. *Review of Economic Studies* 85(2), 1029–1069.
- Doleac, J. L. and L. C. Stein (2013, nov). The Visible Hand: Race and Online Market Outcomes. *The Economic Journal* 123(572), F469–F492.
- Duncan, D. and D. Li (2018). Liar liar: Experimental evidence of the effect of confirmation-reports on dishonesty. *Southern Economic Journal* 84(3), 742–770.
- Edelman, B., M. Luca, and D. Svirsky (2017, apr). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics* 9(2), 1–22.
- Ge, Y., C. Knittel, D. MacKenzie, and S. Zoepf (2016, oct). Racial and Gender Discrimination in Transportation Network Companies. Technical report, National Bureau of Economic Research, Cambridge, MA.
- 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(3), 1219–1260.
- Hedegaard, M. S. and J.-R. Tyran (2018, jan). The Price of Prejudice. *American Economic Journal: Applied Economics* 10(1), 40–63.

- Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics* 14, 399–425.
- Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015a, April). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review* 105(4), 1478–1508.
- Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015b). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review* 105(4), 1478–1508.
- Mason, W. and S. Suri (2011). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavioural Research* 44, 1–23.
- Neumark, D. (2018, sep). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature* 56(3), 799–866.
- Nunley, J. M., M. F. Owens, and R. S. Howard (2011). The effects of information and competition on racial discrimination: Evidence from a field experiment. *Journal of Economic Behavior & Organization* 80(3), 670–679.
- Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5).
- Pope, D. G. and J. R. Sydnor (2011). What’s in a picture? evidence of discrimination from prosper. com. *Journal of Human resources* 46(1), 53–92.
- Riach, P. A. and J. Rich (2002, nov). Field Experiments of Discrimination in the Market Place*. *The Economic Journal* 112(483), F480–F518.
- Zussman, A. (2013, nov). Ethnic Discrimination: Lessons from the Israeli Online Market for Used Cars. *The Economic Journal* 123(572), F433–F468.

7 Tables and Figures

Figure 1: Experimental Design



Notes: Reported is the flow of the experiment. Subjects are recruited on Amazon’s Mechanical Turk (mTurk) to complete a mileage userfee survey on Qualtrics. Subjects are randomly assigned to a treatment group where they are shown a picture of a hand holding a receipt and asked whether they would like to complete a transcription task. Subjects who respond yes transcribe images and then complete a post-experiment survey. Subjects who respond no complete the post-experiment survey.

Figure 2: Treatment Instructions

Thank you! You have earned \$0.65 for completing my survey.

****below is an optional bonus opportunity****

I want to know how much I would pay in mileage tax compared to what I pay now for gasoline tax. To help me, I would like you to transcribe information from my gasoline receipts; this will allow me to estimate my annual gasoline taxes.

Each receipt should take approximately 30 seconds to transcribe, and I will pay you a bonus of \$0.06 for every receipt that you transcribe. You can stop at anytime.

I have included a sample of one of my receipts above. I would like you to transcribe the following information:

- 1.Name of the gas station*
- 2.Date of the purchase*
- 3.Gallons of gasoline purchased*
- 4.Price per gallon*
- 5.Total sale price*

Figure 3: Treatment Pictures



Notes: Reported are the pictures used in the treatment stage of the experiment. The pictures have been compressed significantly to fit side-by-side on one page.

Figure 4: Treatment Question

Would you like to transcribe my gasoline receipts (there is no penalty for opting out of this bonus task)?

Yes

No

CONTINUE

Figure 5: Treatment Task

Purchase information

Gallons of gasoline
 Price per gallon
 Total sale

Receipt information

Store

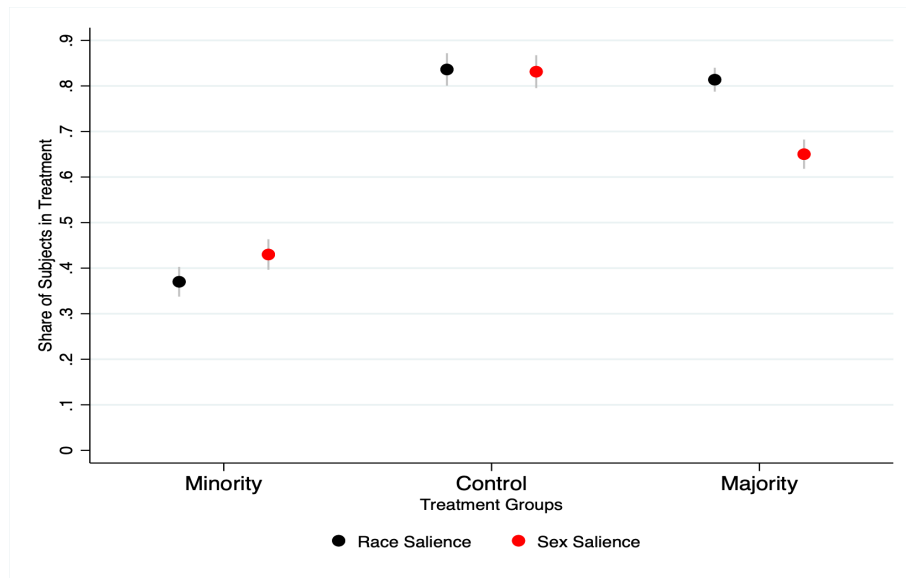
Month

Day

Year

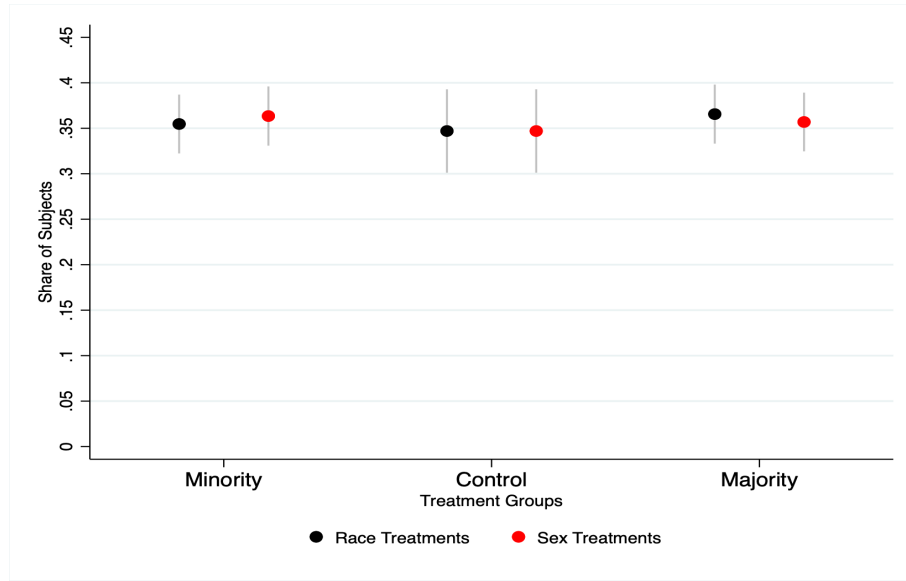
Submit Picture

Figure 6: Salience of race and sex



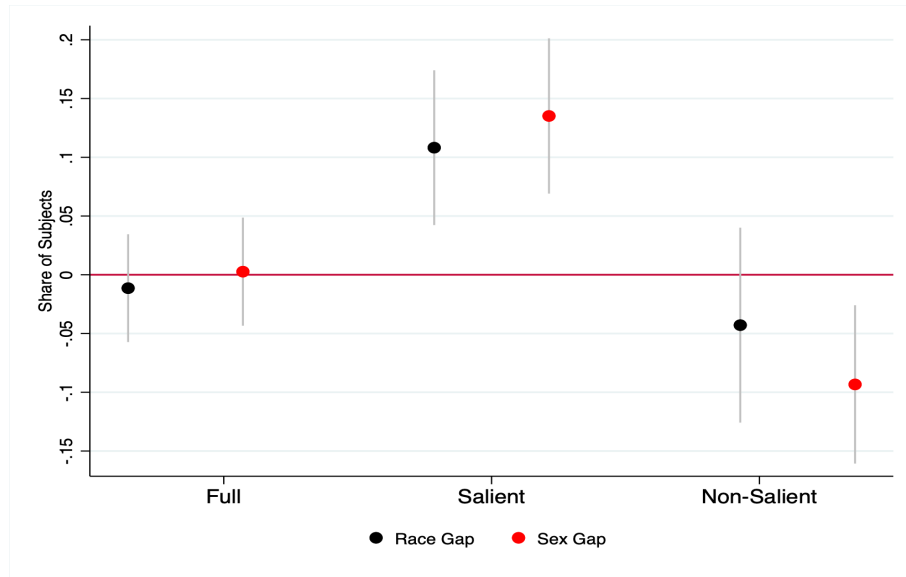
Notes: Reported is the share of subjects whose self-reported perceived treatment matches the actual treatment they are assigned to for race and sex, respectively, along with 95% confidence intervals. Minority refers to Black-hand and female-hand treatments, while majority refers to white-hand and male-hand treatments.

Figure 7: Acceptance Share: by Treatment Group



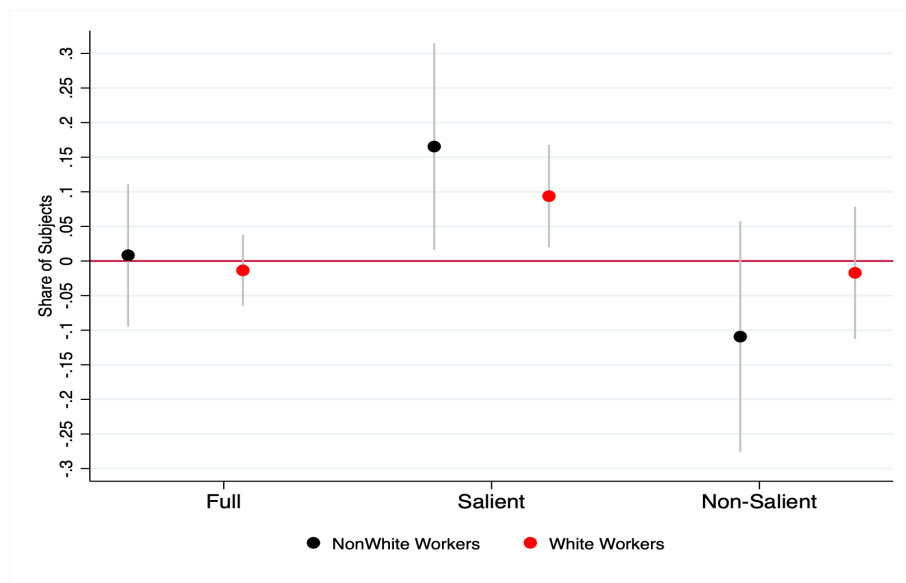
Notes: Reported is the acceptance share by treatment group for race and sex, along with 95% confidence intervals. Acceptance share refers to the share of subjects who agreed to transcribe receipts. Minority refers to Black-hand and female-hand treatments, while majority refers to white-hand and male-hand treatments.

Figure 8: Treatment effect of race and sex on acceptance share



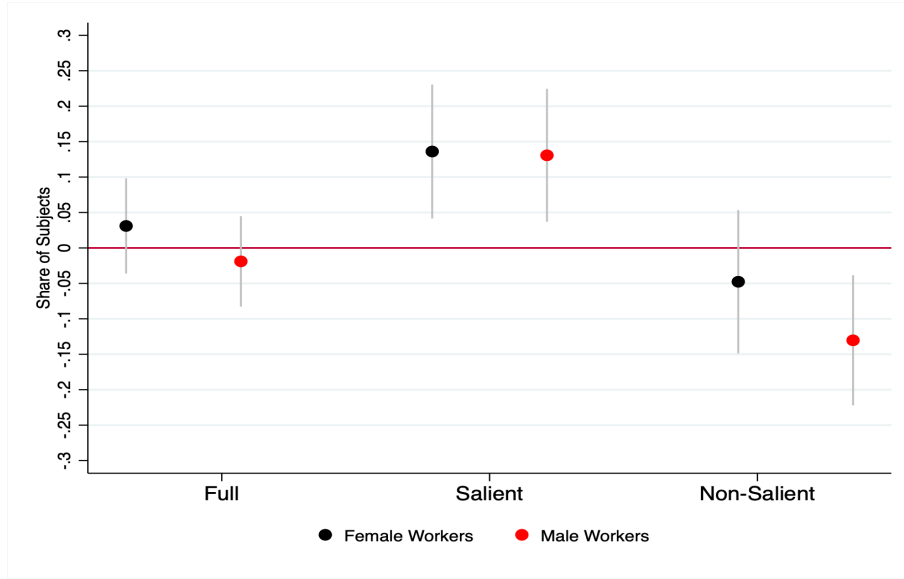
Notes: Reported is the race-gap and sex-gap (with 95% confidence intervals) in the share of subjects who agreed to transcribe pictures among the full sample (Full) and two salience samples. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as difference in acceptance share between the black-hand treatment and the white-hand treatment. The sex-gap is defined as difference in acceptance share between the female-hand treatment and the male-hand treatment.

Figure 9: Treatment effect of race on acceptance share



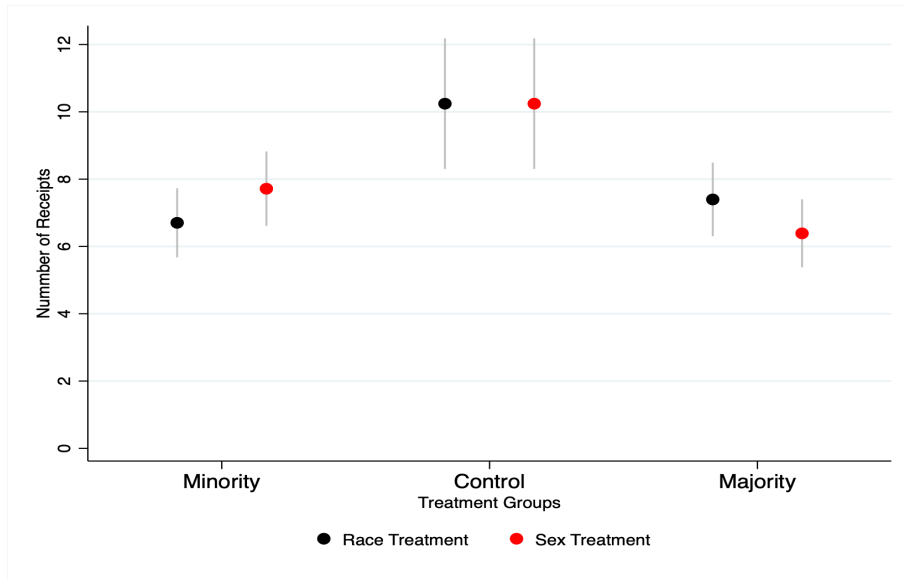
Notes: Reported is the within-group racial-gap (with 95% confidence intervals) in the share of subjects who agreed to transcribe pictures among the full sample (Full) and two salience samples. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as difference in acceptance share between the black-hand treatment and the white-hand treatment.

Figure 10: Treatment effect of sex on acceptance share



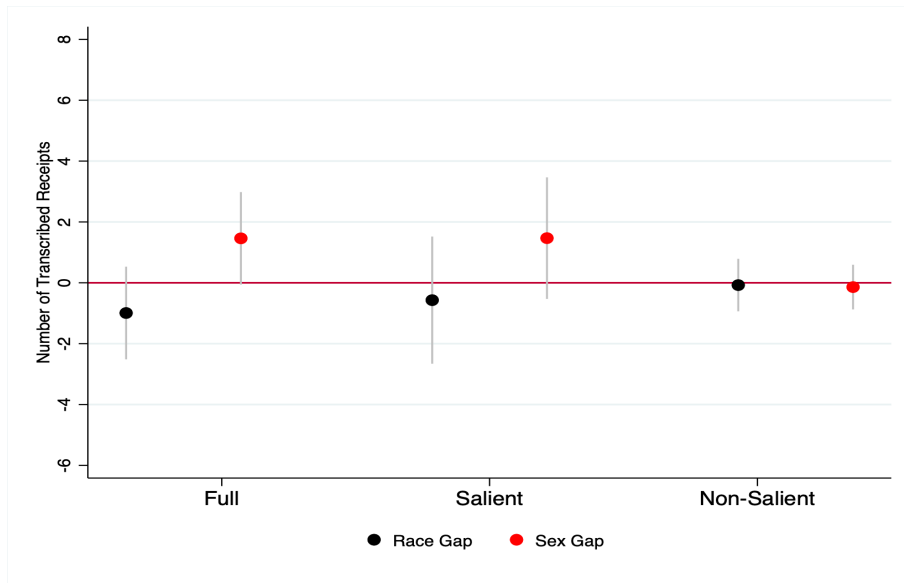
Notes: Reported is the within-group sex-gap (with 95% confidence intervals) in the share of subjects who agreed to transcribe pictures among the full sample (Full) and two salience samples. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The sex-gap is defined as difference in acceptance share between the female-hand treatment and the male-hand treatment.

Figure 11: Mean number of transcribed receipts: by Treatment Group



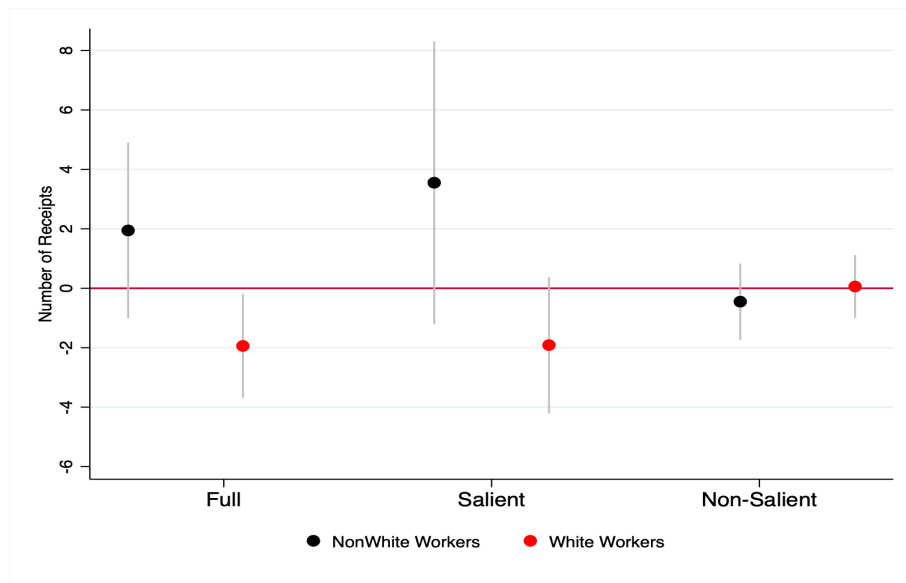
Notes: Reported is the mean number of receipts transcribed by treatment group for race and sex, along with 95% confidence intervals. Minority refers to Black-hand and female-hand treatments, while majority refers to white-hand and male-hand treatments.

Figure 12: Treatment effect of race and sex on number of transcribed receipts



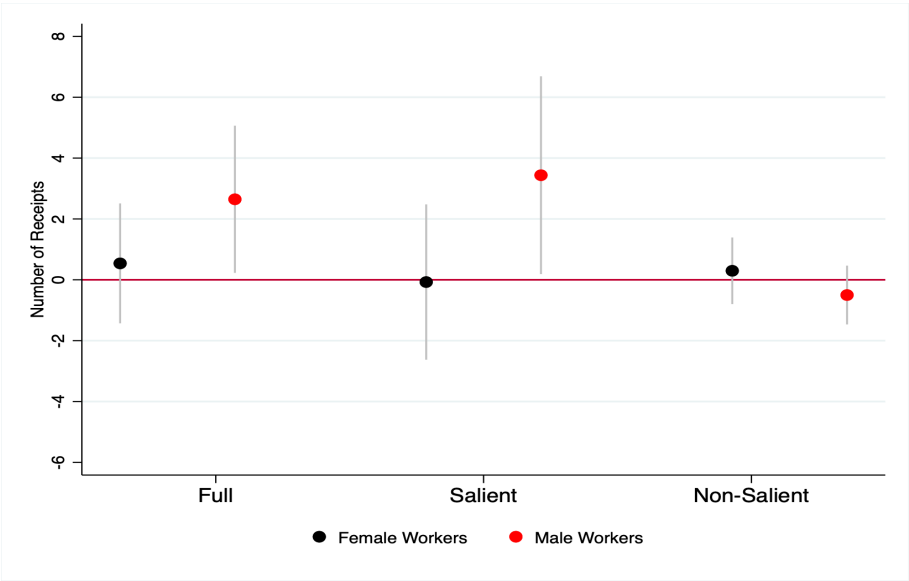
Notes: Reported is the race-gap and sex-gap (with 95% confidence intervals) in the mean number of receipts transcribed by subjects in the full sample (Full) and two salience samples who transcribed at least one receipt. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as difference in acceptance share between the black-hand treatment and the white-hand treatment. The sex-gap is defined as difference in acceptance share between the female-hand treatment and the male-hand treatment.

Figure 13: Treatment effect of race on number of transcribed receipts



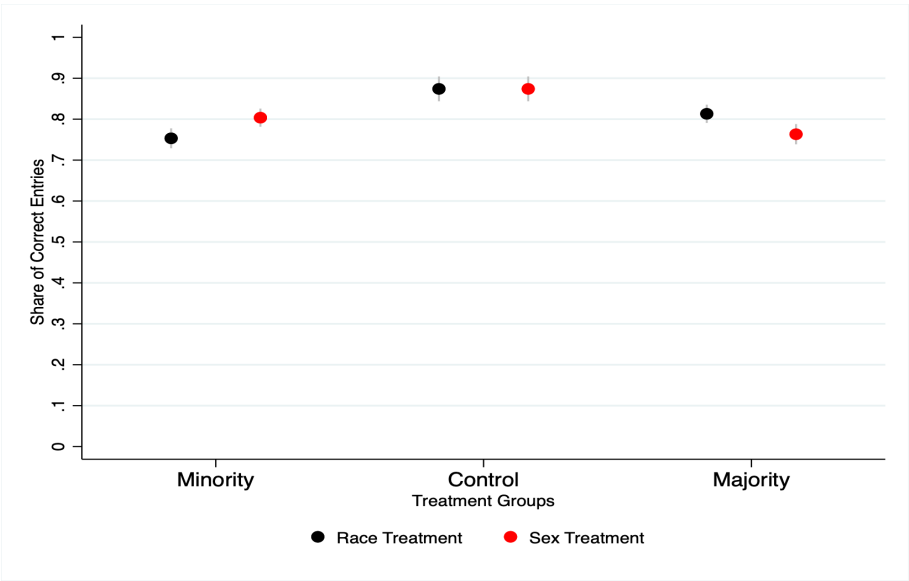
Notes: Reported is the within-group race-gap (with 95% confidence intervals) in the mean number of receipts transcribed by subjects in the full sample (Full) and two salience samples who transcribed at least one receipt. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as difference in acceptance share between the black-hand treatment and the white-hand treatment.

Figure 14: Treatment effect of sex on number of transcribed receipts



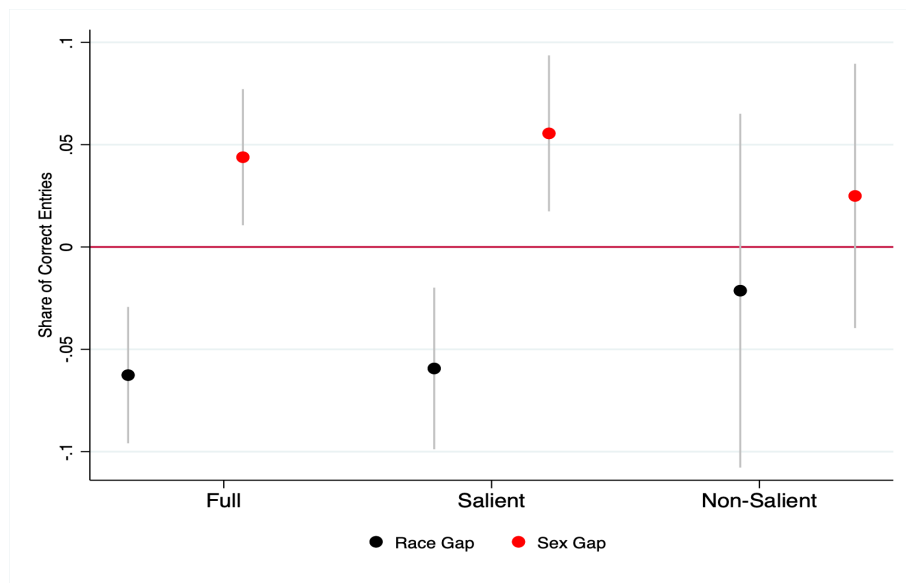
Notes: Reported is the within-group sex-gap (with 95% confidence intervals) in the mean number of receipts transcribed by subjects in the full sample (Full) and two salience samples who transcribed at least one receipt. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The sex-gap is defined as difference in acceptance share between the female-hand treatment and the male-hand treatment.

Figure 15: Accuracy rate by race and sex



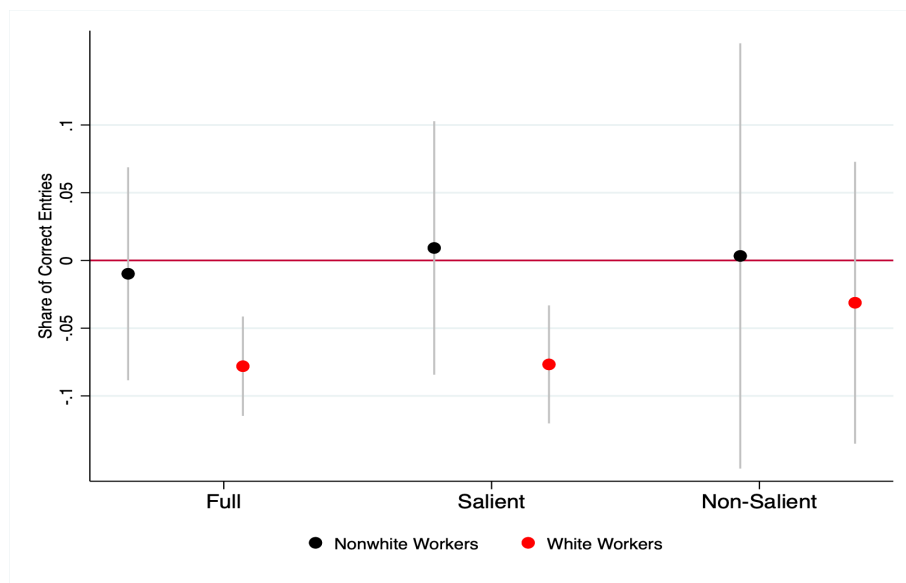
Notes: Reported is the share of accurate transcriptions across treatment groups, along with 95% confidence intervals. Minority refers to Black-hand and female-hand treatments, while majority refers to white-hand and male-hand treatments.

Figure 16: Treatment effect of race and sex on on accuracy rate:



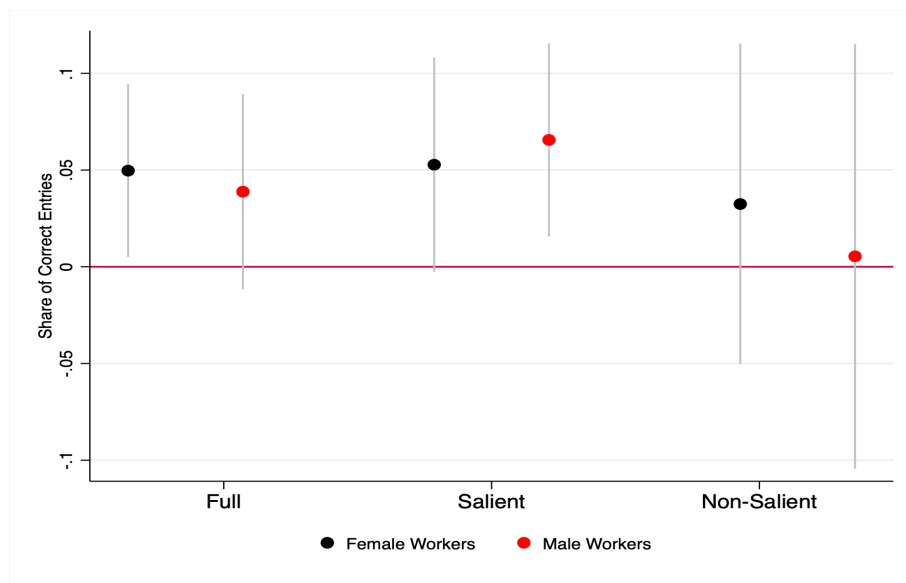
Notes: Reported is the racial-gap and sex-gap (with 95% confidence intervals) in the share of correct entries by subjects in the full sample (Full) and two salience samples who transcribed at least one receipt. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as the difference in share of correct entries between the black-hand treatment and the white-hand treatment. The sex-gap is defined as the difference in share of correct entries between the female-hand treatment and the male-hand treatment.

Figure 17: Treatment effect of race on on accuracy rate:



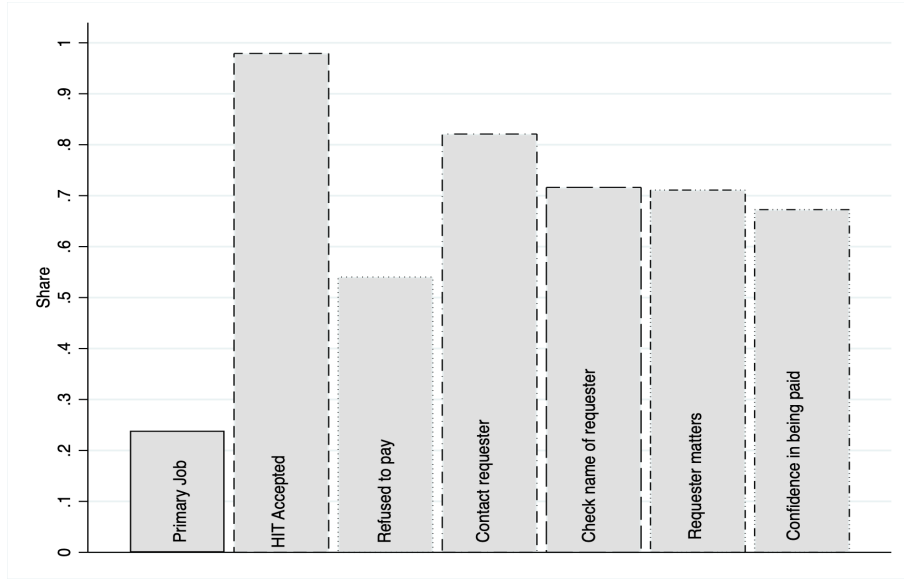
Notes: Reported is the within-group racial-gap (with 95% confidence intervals) in the share of correct entries by subjects in the full sample (Full) and two salience samples who transcribed at least one receipt. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as the difference in share of correct entries between the black-hand treatment and the white-hand treatment.

Figure 18: Treatment effect of sex on on accuracy rate:



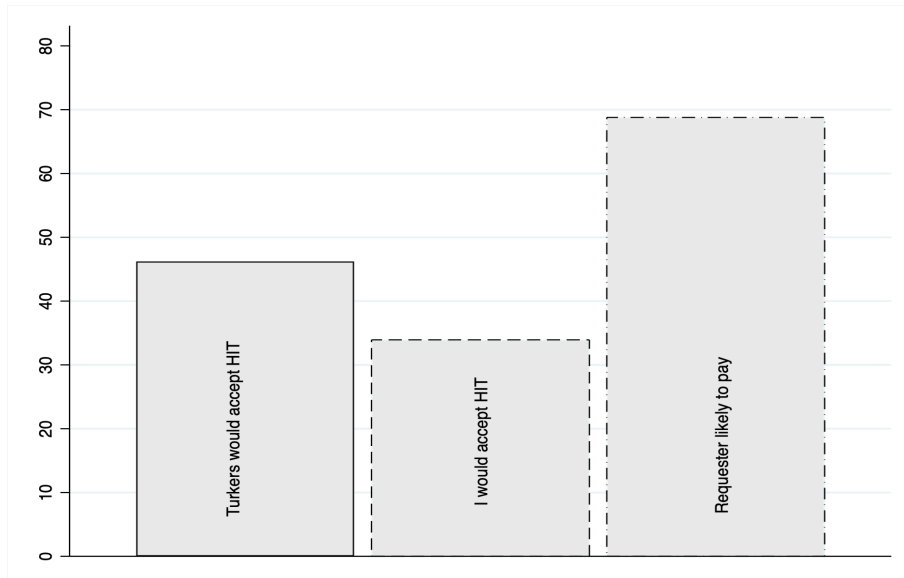
Notes: Reported is the within-group sex-gap (with 95% confidence intervals) in the share of correct entries by subjects in the full sample (Full) and two salience samples who transcribed at least one receipt. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The sex-gap is defined as the difference in share of correct entries between the female-hand treatment and the male-hand treatment.

Figure 19: mTurk Experience Survey



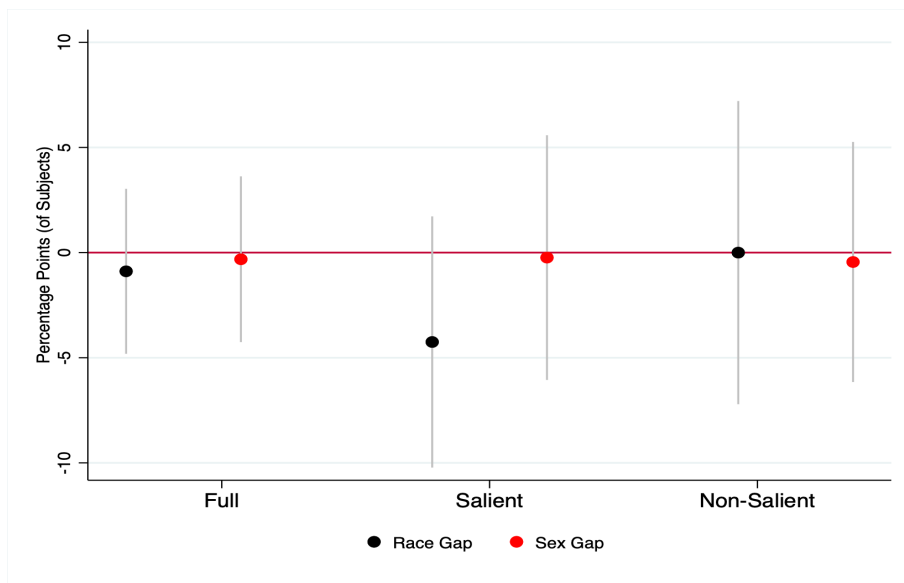
Notes: Reported is the share of mTurkers who answered “yes” to questions regarding their experiences as mTurk workers, including “*Is mTurk your primary job?*”, “*Is your work often accepted?*”, “*Have you experienced requesters’ refusal to pay for work completed?*”, “*Have you contacted requesters before?*”, “*Do you often pay attention to the requester’s name?*”, and “*Will a requester’s characteristics affect your likelihood of accepting a HIT?*”. the question “*how confident are you that a requester will pay for a bonus task*” is measured a scale from 0 to 100. The variable was transformed to a 0 to 1 scale and mean confidence level is reported.

Figure 20: mTurk Experience Survey



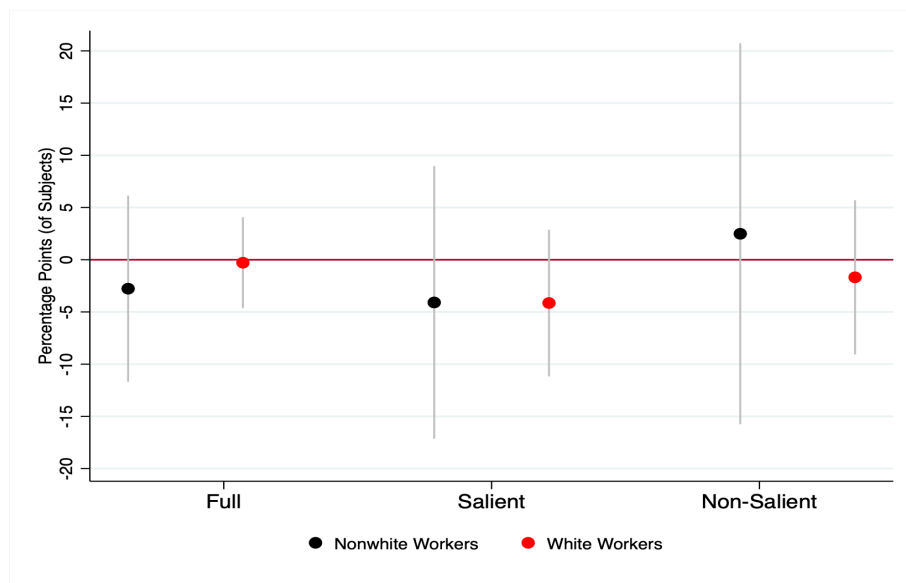
Notes: Reported is subjects’ belief about the percentage of other mTurkers who would accept the hypothetical task, the percentage of subjects who would accept the task themselves, and subjects’ confidence level that the requester of the task would pay the stated bonus.

Figure 21: Requester will pay bonus - Treatment Effects



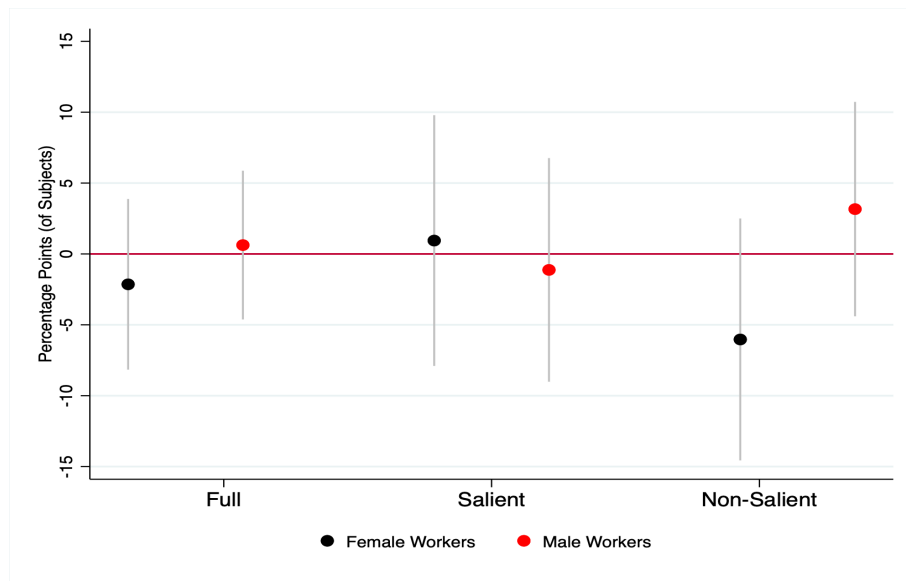
Notes: Reported is the racial-gap and sex-gap (with 95% confidence intervals) in subjects' confidence level that the requester of the task would pay the stated bonus by subjects in the full sample (Full) and two salience samples. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as the difference in share of correct entries between the black-hand treatment and the white-hand treatment. The sex-gap is defined as the difference in share of correct entries between the female-hand treatment and the male-hand treatment.

Figure 22: Requester will pay bonus - Treatment Effects by Subjects' Race



Notes: Reported is the within-group racial-gap (with 95% confidence intervals) in subjects' confidence level that the requester of the task would pay the stated bonus by subjects in the full sample (Full) and two salience samples. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The racial-gap is defined as the difference in share of correct entries between the black-hand treatment and the white-hand treatment.

Figure 23: Requester will pay bonus - Treatment Effects by Subjects' Sex



Notes: Reported is the within-group sex-gap (with 95% confidence intervals) in subjects' confidence level that the requester of the task would pay the stated bonus by subjects in the full sample (Full) and two salience samples. Salient includes only subjects who correctly identified the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications and Non-Salient includes subjects who did not correctly identify the race of the employer for the race-gap specifications and sex of the employer for the sex-gap specifications. The sex-gap is defined as the difference in perception between the female-hand treatments and male-hand treatments.

Table 1: Summary Statistics of Covariates

Variables	Black	White	No Pic	Female	Male	Total
Age	37.46	37.49	36.64	37.80	37.15	37.31
White	0.80	0.77	0.76	0.79	0.79	0.78
Sex	0.50	0.49	0.43	0.51	0.48	0.48
Urban	0.50	0.54	0.49	0.53	0.51	0.52
Duration (mins)	7.11	7.48	7.68	7.25	7.34	7.37
High School	0.11	0.09	0.11	0.08	0.11	0.10
Some College	0.25	0.22	0.23	0.25	0.22	0.23
2-Year College	0.11	0.12	0.10	0.12	0.12	0.11
B.Sc.	0.38	0.42	0.41	0.38	0.42	0.40
Graduate	0.16	0.15	0.16	0.18	0.13	0.16
N. Obs.	843	848	415	842	849	2106

Notes: Reported is the mean of each variable by treatment group. We combine data by race and sex. ‘No Pic’ is the control group.

Table 2: Balancedness test

Demographic	Black v Control		White vs Control		Female V Control		Male V Control		W vs B		F vs M	
	Difference	Pvalue	Difference	Pvalue	Difference	Pvalue	Difference	Pvalue	Pvalue	Pvalue	Pvalue	Pvalue
Age	0.825	0.378	0.850	0.270	1.163	0.120	0.515	0.664	0.834	0.203		
White	0.048	0.052	0.011	0.664	0.028	0.256	0.030	0.226	0.067	0.929		
Sex	0.069	0.021	0.053	0.075	0.072	0.016	0.050	0.092	0.512	0.369		
Urban	0.012	0.702	0.049	0.104	0.041	0.170	0.019	0.522	0.126	0.365		
Duration (Mins)	-0.572	0.943	-0.203	0.920	-0.430	0.704	-0.344	0.685	0.780	0.340		
High School	0.001	0.968	-0.019	0.282	-0.024	0.161	0.006	0.754	0.176	0.038		
Some College	0.017	0.516	-0.007	0.773	0.018	0.480	-0.009	0.729	0.247	0.193		
2-Year College	0.009	0.623	0.023	0.239	0.015	0.421	0.017	0.381	0.388	0.929		
B.Sc.	-0.030	0.304	0.009	0.759	-0.031	0.293	0.010	0.742	0.101	0.089		
Graduate	0.004	0.873	-0.006	0.792	0.022	0.341	-0.024	0.259	0.602	0.011		

Notes: Reported is the difference in mean between treatment and control groups for each variable. Pvalues from a ranksum test of the differences in means between treatment groups are also reported. W vs B is white compared to black, and F vs M is female compared to male.

Table 3: Characteristics of Salient and Non-salient samples

	Race		Sex		Full sample
	Non-Salient	Salient	Non-Salient	Salient	
Age	39.16	36.28	38.52	36.50	37.31
White	0.78	0.78	0.78	0.78	0.78
Sex	0.49	0.48	0.49	0.48	0.48
Urban	0.50	0.53	0.51	0.52	0.52
Duration	6.42	7.91	6.40	8.03	7.37
High School	0.11	0.09	0.10	0.10	0.10
Some College	0.20	0.25	0.23	0.24	0.23
2-Year College	0.12	0.11	0.12	0.11	0.11
B.Sc.	0.39	0.40	0.37	0.42	0.40
Graduate	0.18	0.14	0.18	0.14	0.16

Notes: Reported is the mean of each variable for the salient and non-salient samples. A subject is in the race-salient sample if her self-reported perceived race treatment matches her assigned race treatment and the non-salient sample otherwise. A subject is in the sex-salient sample if her self-reported perceived sex treatment matches her assigned sex treatment and the non-salient sample otherwise.

Table 4: Summary statistics: mTurk worker experience survey

Variables	Black	White	No Pic	Female	Male
Age Group:					
18 to 24	0.13	0.12	0.13	0.14	0.10
25 to 34	0.45	0.42	0.44	0.40	0.48
35 to 44	0.22	0.26	0.27	0.24	0.24
45 to 54	0.13	0.12	0.10	0.13	0.13
55 to 64	0.05	0.06	0.06	0.06	0.05
>65	0.01	0.02	0.01	0.02	0.01
White	0.76	0.76	0.72	0.76	0.77
Sex	0.45	0.48	0.49	0.46	0.47
Duration (mins)	4.65	4.60	4.39	4.57	4.68
High School	0.10	0.12	0.10	0.12	0.10
Some College	0.27	0.21	0.23	0.23	0.24
2-Year College	0.12	0.13	0.11	0.12	0.13
B.Sc.	0.38	0.39	0.42	0.38	0.39
Graduate	0.13	0.16	0.14	0.15	0.14
N. Obs.	347	353	166	348	352

Notes: Reported is the mean of each variable from the mTurk worker experience survey. We combine data by race and sex. 'No Pic' is the control group.

Table 5: Balancedness test: mTurk worker experience survey

Variables	Male v Control	Female control	White vs Control	Black vs Control	Female v Male	White vs Black
18 to 24	0.411	0.598	0.807	0.993	0.095	0.753
25 to 34	0.425	0.386	0.750	0.787	0.038	0.463
35 to 44	0.426	0.513	0.748	0.251	0.862	0.308
45 to 54	0.343	0.281	0.395	0.240	0.864	0.670
55 to 64	0.569	0.896	0.973	0.697	0.390	0.661
≥65	0.762	0.228	0.313	0.553	0.197	0.542
White	0.246	0.425	0.302	0.355	0.660	0.899
Sex	0.592	0.468	0.701	0.378	0.812	0.534
Duration	0.712	0.885	0.949	0.870	0.537	0.830
High School	0.994	0.416	0.502	0.874	0.306	0.516
Some College	0.816	0.956	0.564	0.464	0.720	0.102
2-Year College	0.603	0.912	0.674	0.830	0.606	0.796
B.Sc.	0.525	0.468	0.509	0.483	0.910	0.957
Graduate	0.915	0.809	0.553	0.782	0.865	0.276

Notes: Reported is the difference in mean between treatment and control groups for each variable. Pvalues from a ranksum test of the differences in means between treatment groups are also reported. W vs B is white compared to black, and F vs M is female compared to male.

Table 6: Saliency of Treatment in Experiment and Survey

Treatment	Race Salient		Sex Salient	
	Experiment	Survey	Experiment	Survey
BF	0.43	0.41	0.39	0.41
	419	185	419	185
BM	0.31	0.32	0.62	0.55
	424	192	424	192
Control	0.84	0.88	0.83	0.88
	415	190	415	190
WF	0.84	0.84	0.47	0.47
	423	200	423	200
WM	0.79	0.78	0.68	0.62
	425	188	425	188
Total	0.64	0.65	0.60	0.59
	2106	955	2106	955

Notes: Reported is the share of subjects whose self-reported perceived treatment matches the actual treatment they are assigned for race and sex, respectively, in the real-effort experiment and the mTurk user-survey. A subject is in the race-salient sample if her self-reported perceived race treatment matches her assigned race treatment and the non-salient sample otherwise. A subject is in the sex-salient sample if her self-reported perceived sex treatment matches her assigned sex treatment and the non-salient sample otherwise. BF includes subjects who were assigned to the black female hand treatment; BM includes subjects who were assigned to the black male hand treatment; Control includes subjects who were assigned to the control group; WF includes subjects who were assigned to the white female hand treatment; and WM includes subjects who were assigned to the white male hand treatment.

8 Appendix

Table 7: Summary Statistics by Treatment Group

Variables	BF	BM	Control	WF	WM	Total
Age	38.224	36.712	36.639	37.383	37.594	37.311
White	0.809	0.800	0.757	0.761	0.774	0.780
Sex	0.516	0.491	0.434	0.496	0.478	0.483
Urban	0.536	0.470	0.491	0.530	0.551	0.516
Duration (mins)	6.793	7.429	7.684	7.712	7.252	7.374
High School	0.081	0.132	0.106	0.083	0.092	0.099
Some College	0.265	0.226	0.229	0.229	0.214	0.233
2-Year College	0.117	0.104	0.101	0.116	0.132	0.114
B.Sc.	0.346	0.408	0.407	0.407	0.426	0.399
Graduate	0.191	0.130	0.157	0.165	0.136	0.156
N. Obs.	419	424	415	423	425	2106

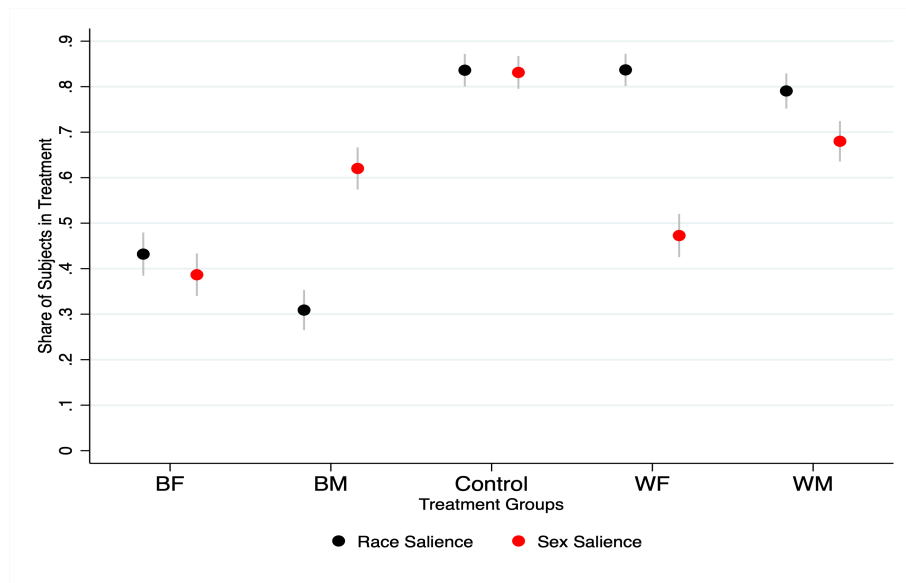
Notes: Reported is the mean of each variable by treatment group. BF is black female hand, BM is black male, WF is white female and WM is white male.

Table 8: Summary Statistics for mTurk worker experience survey

Variables	BF	BM	Control	WF	WM	Total
18 to 24	0.17	0.09	0.13	0.12	0.11	0.12
25 to 34	0.36	0.54	0.44	0.44	0.41	0.44
35 to 44	0.24	0.21	0.27	0.25	0.27	0.25
45 to 54	0.15	0.11	0.10	0.11	0.14	0.12
55 to 64	0.06	0.04	0.06	0.07	0.05	0.06
>65	0.02	0.01	0.01	0.02	0.01	0.01
White	0.75	0.77	0.72	0.76	0.77	0.76
Sex	0.41	0.49	0.49	0.51	0.44	0.47
Duration	4.78	4.53	4.39	4.38	4.83	4.58
High School	0.10	0.10	0.10	0.14	0.09	0.11
Some College	0.27	0.26	0.23	0.20	0.22	0.24
2-Year College	0.10	0.14	0.11	0.13	0.12	0.12
B.Sc.	0.41	0.35	0.42	0.35	0.42	0.39
Graduate	0.12	0.14	0.14	0.17	0.14	0.14
	169	178	166	179	174	866

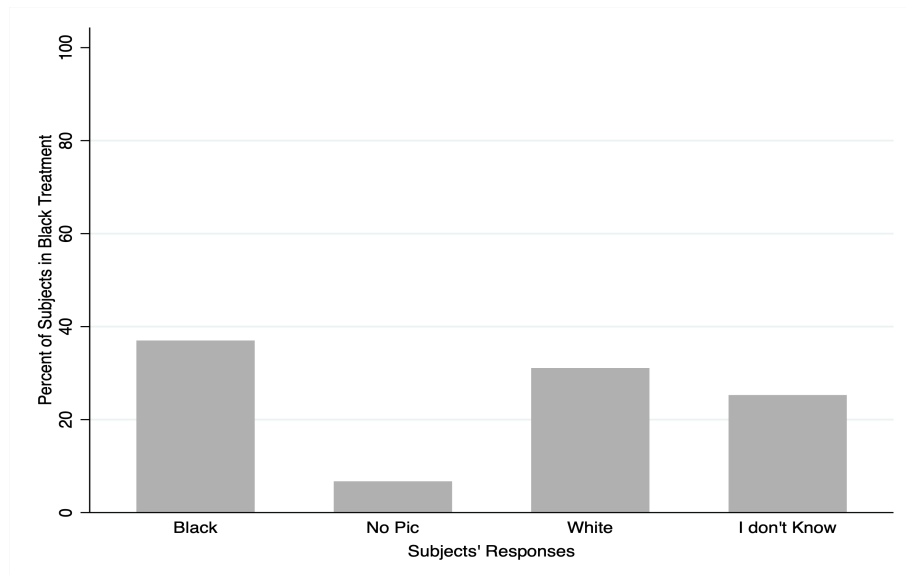
Notes: Reported is the mean of each variable by treatment group. BF is black female hand, BM is black male, WF is white female and WM is white male.

Figure 24: Race salience by detailed treatment group



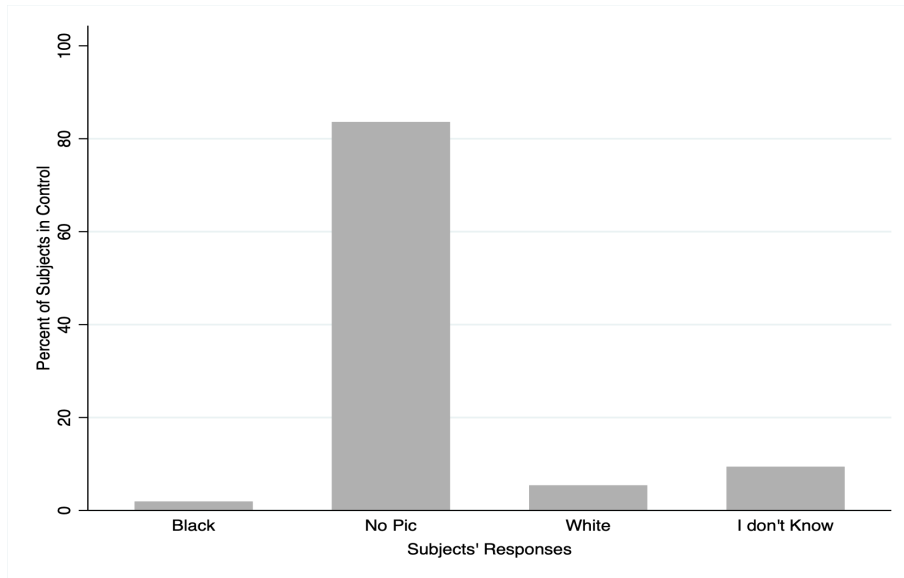
Notes: Reported is the share of race-salient and sex-salient subjects by treatment group. Race-salience includes only subjects who correctly identified the race of the employer and sex-salience includes only subjects who correctly identified the sex of the employer. BF includes subjects who were assigned to the black female hand treatment; BM includes subjects who were assigned to the black male hand treatment; Control includes subjects who were assigned to the control group; WF includes subjects who were assigned to the white female hand treatment; and WM includes subjects who were assigned to the white male hand treatment.

Figure 25: Race Salience in the Black Treatment Group



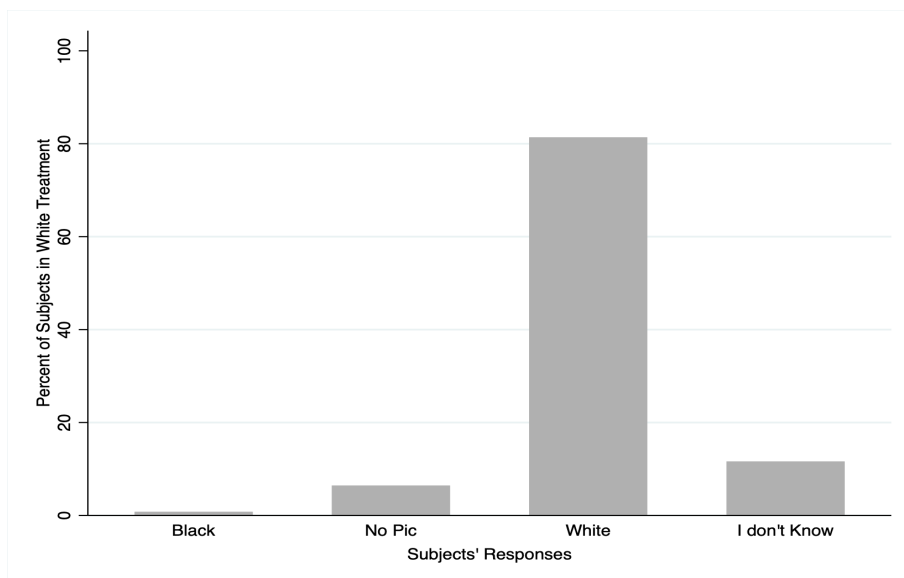
Notes: Reported is the percent of subjects in the black treatment who gave each of the possible responses to the question: "What is the race of the person holding the receipt in the picture?".

Figure 26: Race Salience in the Control Group



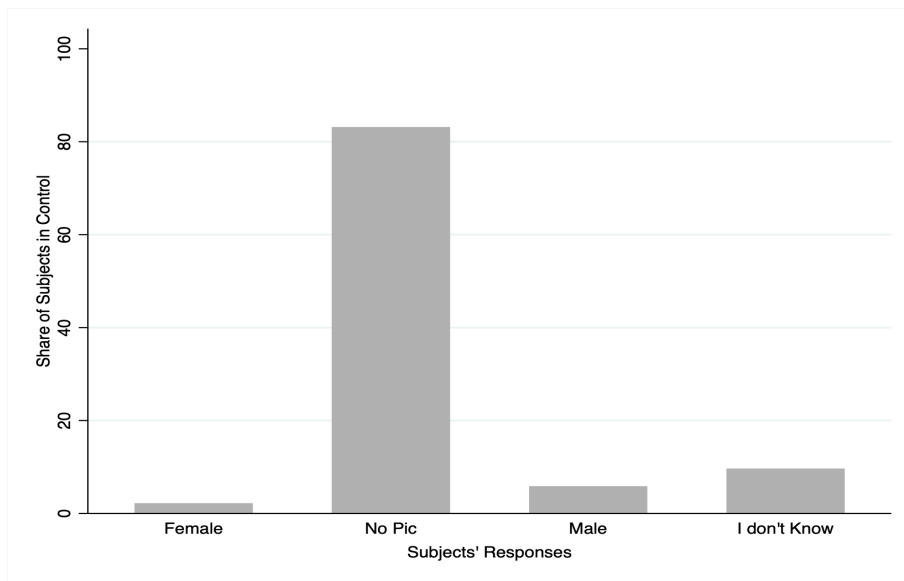
Notes: Reported is the percent of subjects in the control group who gave each of the possible responses to the question: “What is the race of the person holding the receipt in the picture?”.

Figure 27: Race Salience in the White Treatment Group



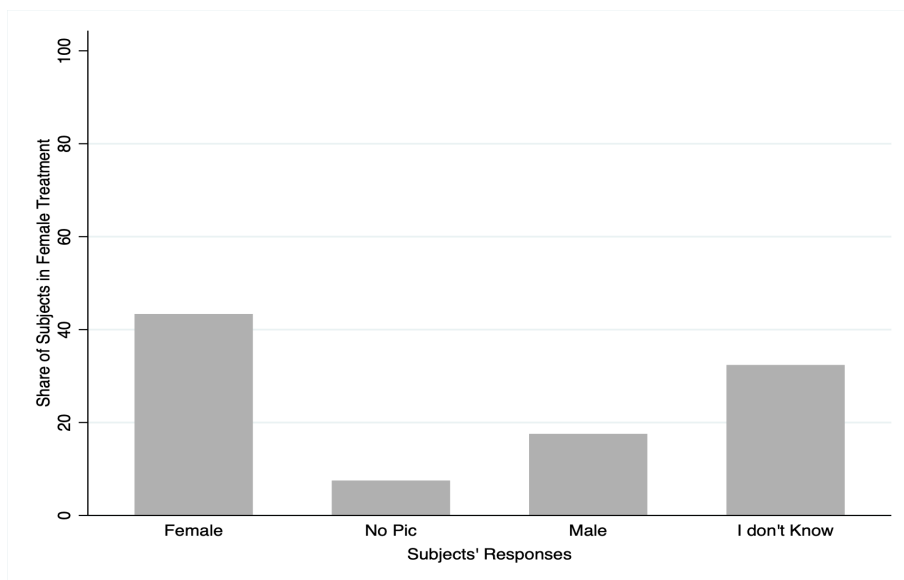
Notes: Reported is the percent of subjects in the white treatment who gave each of the possible responses to the question: “What is the race of the person holding the receipt in the picture?”.

Figure 28: Sex Salience in the Control Group



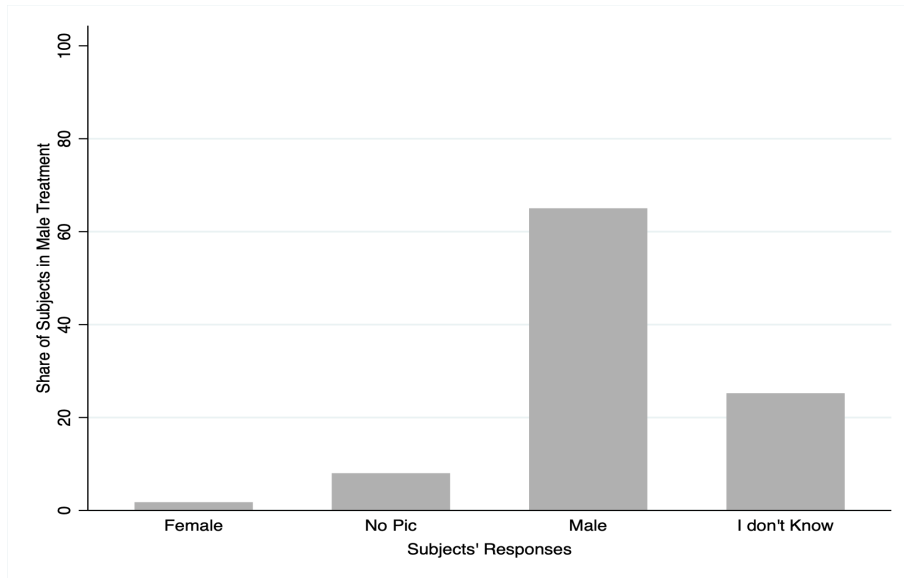
Notes: Reported is the percent of subjects in the control group who gave each of the possible responses to the question: “What is the sex of the person holding the receipt in the picture?”.

Figure 29: Sex Salience in the Female Treatment Group



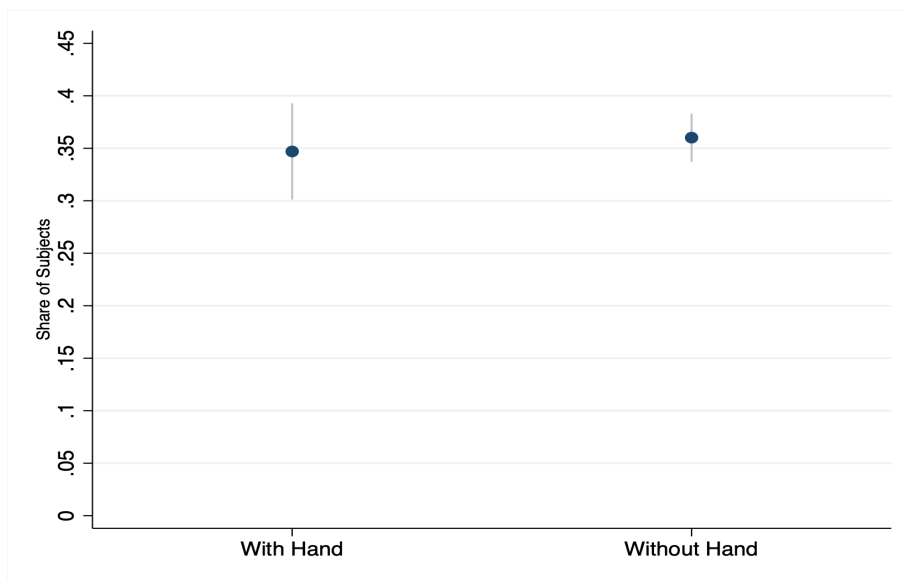
Notes: Reported is the percent of subjects in the female treatment group who gave each of the possible responses to the question: “What is the sex of the person holding the receipt in the picture?”.

Figure 30: Sex Salience in the Male Treatment Group



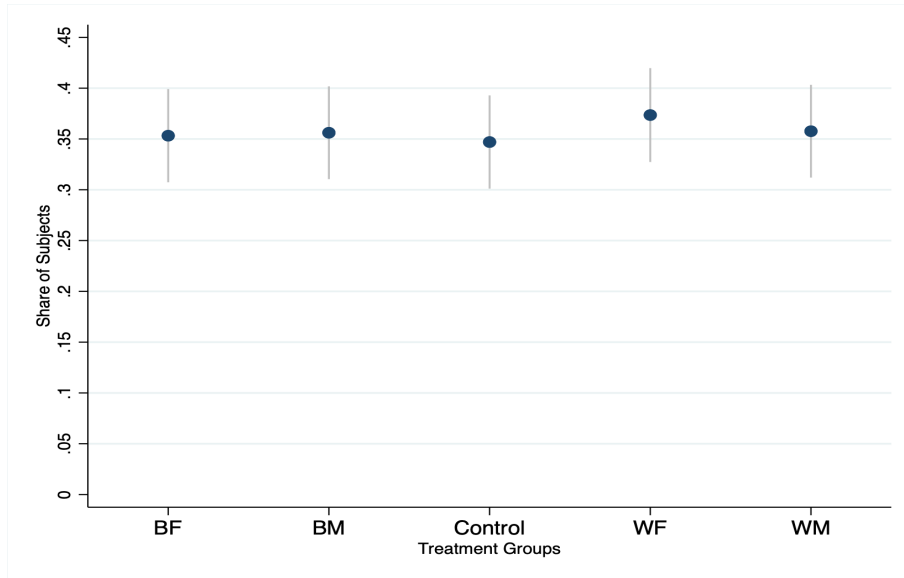
Notes: Reported is the percent of subjects in the male treatment group who gave each of the possible responses to the question: “What is the sex of the person holding the receipt in the picture?”.

Figure 31: Transcription share



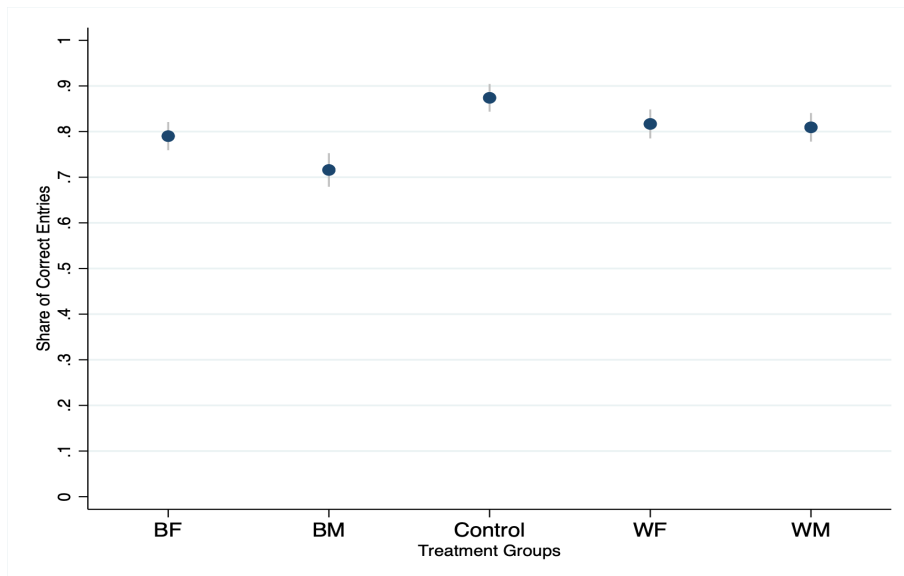
Notes: Reported is the share of subjects who agreed to transcribe receipts. ‘With Hand’ indicates that the image of the receipt included a hand; ‘Without Hand’ indicates subjects in the control group.

Figure 32: Transcription share by treatment groups



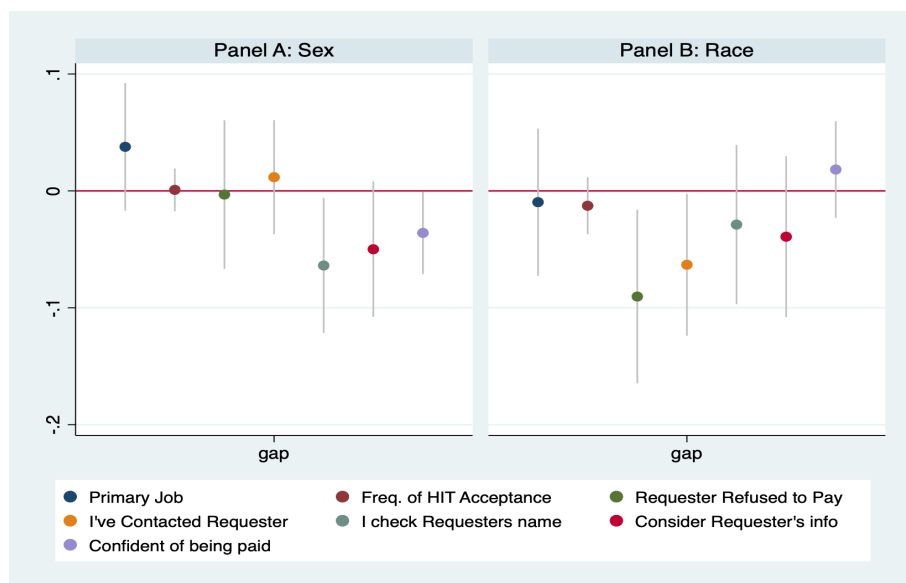
Notes: Reported is the share of subjects who agreed to transcribe receipts in each treatment group. BF includes subjects who were assigned to the black female hand treatment; BM includes subjects who were assigned to the black male hand treatment; Control includes subjects who were assigned to the control group; WF includes subjects who were assigned to the white female hand treatment; and WM includes subjects who were assigned to the white male hand treatment.

Figure 33: Accuracy Rate by Treatment Group



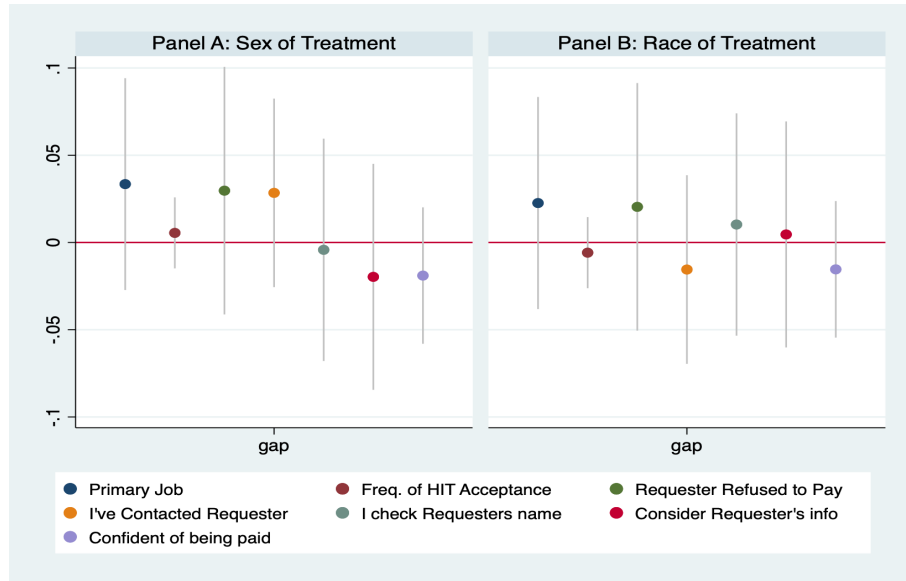
Notes: Reported is the share of correct entries in each treatment group. BF includes subjects who were assigned to the black female hand treatment; BM includes subjects who were assigned to the black male hand treatment; Control includes subjects who were assigned to the control group; WF includes subjects who were assigned to the white female hand treatment; and WM includes subjects who were assigned to the white male hand treatment.

Figure 34: mTurker Work Experience Survey



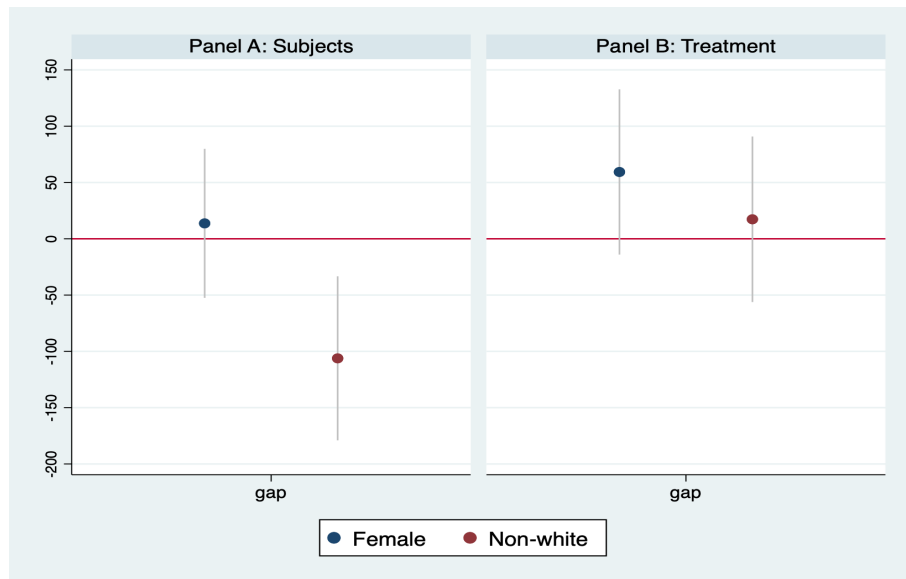
Notes: Panel A reports the difference (between female and male subjects) in the share of mTurkers who answered yes to questions regarding their experiences as mTurk workers and 95% confidence intervals. Panel B is similar to Panel A except that differences are calculated between non-white and white subjects. Questions include the following: *Is mTurk your primary job?*, *“Is your work often accepted?”*, *“Have you experienced requesters’ refusal to pay for work completed?”*, *“Have you contacted requesters before?”*, *“Do you often pay attention to the requester’s name?”*, and *“Will a requester’s characteristics affect your likelihood of accepting a HIT?”*. The question *“how confident are you that a requester will pay for a bonus task?”* is measured a scale from 0 to 100. The variable was transformed to a 0 to 1 scale.

Figure 35: mTurker Work Experience Survey



Notes: Panel A reports the difference (between female and male subjects) in the share of mTurkers who answered “yes” to questions regarding their experiences as mTurk workers and 95% confidence intervals. Panel B is similar to Panel A except that differences are calculated between black and white treatments. Questions include the following: “*Is mTurk your primary job?*”, “*Is your work often accepted?*”, “*Have you experienced requesters’ refusal to pay for work completed?*”, “*Have you contacted requesters before?*”, “*Do you often pay attention to the requester’s name?*”, and “*Will a requester’s characteristics affect your likelihood of accepting a HIT?*”. The question “*how confident are you that a requester will pay for a bonus task?*” is measured a scale from 0 to 100. The variable was transformed to a 0 to 1 scale.

Figure 36: mTurker Work Experience Survey



Notes: Reported is the gap in the mean number of monthly HITs completed by subjects along with 95% confidence intervals. Differences are calculated between female and male subjects, and between non-white and white subjects in Panel A. Differences are calculated between female and male treatments and between black and white treatments in Panel B.