STEREOTYPE-CONFIRMING RACE DISCRIMINATION ACROSS MULTIPLE MINORITIES[†]

HYOEUN PARK[*]

ABSTRACT. In this paper, we study race discrimination with multiple minorities using a resume evaluation experiment. Studies with only one minority group might miss the different types of discrimination that different groups might face. Our result shows that managers in the lab do not have accurate beliefs regarding the performance of workers with racial information. However, managers' inaccurate beliefs do not favor any certain race group. Rather, that is due to miscalibration that gives a favor to low-performing groups. Also, we show that racial information affects wages. Racial information barely reduces wages for Whites, while Asians and Black people often get paid less with their racial information. This is evidence of taste-based discrimination. More importantly, Asians and Black people are discriminated against with different patterns. Asians with high signals are discriminated against, and Black people with low signals are discriminated against. Thus, discrimination occurs to people who conform to existing stereotypes.

Keywords: Race discrimination, Laboratory experiment, Resume evaluation experiment

JEL Classification: C91; J71.

Draft: January 1, 2024
Latest draft available Here

[†] The author thanks Paul Healy, Yaron Azrieli, Trevon Logan, John Rehbeck, Kirby Nielsen, Lucas Coffman, Sally Sadoff, Tim Salmon, participants at the ESA Global Meetings (2023), 2023 North American ESA Meeting, and members of the OSU Theory/Experimental Reading Group for valuable comments. Funding for this project was provided by the Decision Science Collaborative Research Grant at OSU. This project was determined exempt from IRB review under protocol 2022E1146.

[*] Dept. of Economics, Ohio State University, Columbus, Ohio 43210, U.S.A.; park.2774@osu.edu.

1. INTRODUCTION

Economists have studied discrimination across many different dimensions, such as gender, race, ethnicity, or sexual orientation (Blau and Kahn (2017), Lang and Spitzer (2020), Badgett et al. (2021), Booth et al. (2012)). However, most researchers adopt a binary classification of groups: one majority and one minority. While this simplification allows for clear analyses, it might miss important aspects of discrimination. For example, when studying race discrimination, both Black and Asian groups are minorities, but it is questionable whether the two groups face similar types of discrimination. We know that different races have different stereotypes: Asians are thought to be academically successful but quiet, while people stereotype Black people as athletically and musically inclined but lazy (Harpalani, 2022, Wood and Chesser, 1994, Mayovich, 1972). And one race group might face taste-based discrimination, while another group faces a mixture of taste-based and inaccurate statistical discrimination. In other words, there can be heterogeneity in discrimination faced by different minority groups. If different minority groups face discrimination in different ways, focusing only on one group or one measure of performance (such as academic achievement) may cause an underestimation of discrimination against other minority groups, and may ultimately lead to ineffective or even harmful policy recommendations.

In this paper, we study race discrimination with multiple minority groups. We consider three race groups: Whites (majority), Black people (minority), and Asians (minority).[1] We measure both tasted-based discrimination (Becker, 1971) and statistical discrimination (Phelps, 1972, Arrow, 1974). By performing a controlled laboratory study, we can measure differences in beliefs across races and compare beliefs to objective performance as in Bohren et al. (2019a).

We focus on a labor market setting that involves belief updating and a hiring decision. We employ a laboratory experiment in which subjects acting as managers evaluate

[1]For this study, we focus only on race, but follow-up work could also study discrimination against minority ethnicities, such as Hispanics, or a larger set of racial minorities.

workers' resumes. This is similar to resume audit studies in the fields (Bertrand and Mullainathan, 2004, Kline and Walters, 2021, Kline et al., 2022) but avoids deception since we use actual workers' information. This design is built on the design of Bohren et al. (2019a). In their design, subjects in the role of managers are given quasi-resumes of real workers (who were also subjects) and asked the managers to report (1) their willingness to pay to hire each worker and (2) their beliefs about the worker's performance on a math task. They included multiple dimensions such as gender, nationality, and age, but for each dimension, they considered only a binary classification. We also have real worker subjects and give managers quasi-resumes to evaluate. However, we consider multiple minority groups and multiple skills and have a control treatment without group identity so we can identify the causal effect of race information. Rather than focusing only on math tasks, we introduce two more tasks to encompass diverse aspects of discrimination: a social skills task and a "combined" task that uses both math and social skills. A comparison between social and math skills can capture different dimensions of stereotypes. The combined task may be more analogous to the criteria actually used in hiring decisions.

Subjects in the role of workers complete three tasks, measuring each of the three skills. We also generate resumes for workers using actual data about their demographics, hometown, high school activities, etc., Then, subjects in the role of managers evaluate these resumes, using them to form beliefs about the workers' performance on the three tasks. They report these beliefs about workers' performance and decide on wages that are actually paid to workers. Thus, we can compare workers' real performance to managers' beliefs, and we can see how those beliefs are related to their willingness to pay. For the managers, we had two between-subject treatments: Race-Revealed and Race-Blind. The only difference was whether they were provided with racial information about the workers. This approach allows us to measure directly the impact of race information on managers' decisions.

Our results show that managers in the lab do not have accurate beliefs regarding the performance of workers. These inaccurate beliefs favor certain groups, but our Race-Blind control reveals that this is not because of their race. Rather, managers' miscalibrated beliefs favor low-performing groups, even when no race information is given. Also, we show that race information affects wages. Specifically, race information does not reduce wages for Whites, while Asians and Black people often experience lower wages when their race is revealed. This is direct evidence of taste-based discrimination. More importantly, Asians and Black people face different patterns of discrimination. Managers discriminate against Asians whose high school activities signal high performance, while they discriminate against Black people whose resumes signal low performance. This pattern is consistent with existing stereotypes. Common stereotypes are that Asians are academically successful while Black people are not (Harpalani, 2022, Wood and Chesser, 1994). Thus, the pattern we found implies that if a minority candidate fits to stereotype of their racial group, then they are discriminated against.

In an additional experiment, we used a method called Item Counting Technique (ICT, (Miller, 1984)) to examine whether people are willing to fix discrimination. ICT is used when a researcher wants to elicit responses to socially sensitive questions but worries their answers will be skewed by social desirability biases. Rather than directly asking a sensitive question to respondents, a researcher creates a set of statements. In one treatment, they include the sensitive question in the set, while in another treatment they do not. Then, they ask respondents how many of the statements they agree with and compare distributions from the two treatments. A difference in the two distributions indicates a hidden preference for the sensitivity question of interest. We follow Coffman et al. (2017) and use a modified version of ICT. Our results show that a substantial portion of people are not willing to fix discrimination in a workplace setting. And, perhaps surprisingly, those who do not want to fix discrimination are willing to admit this when asked directly.

In summary, we document that different minority groups do, in fact, experience different forms of discrimination in different scenarios. Therefore, there might be no one-size-fits-all policy that could fix all discrimination for all minority groups. This implication also casts doubt on the recent decision of the Supreme Court that bans affirmative action in college admissions. The logic is that affirmative action, which gives a minimum quota to Black and Hispanic students, would harm Asian students. However, our results suggest that Black people are discriminated against when their resume signals low performance. This could be similar to candidates who are marginal in an admission decision, so they could be underrepresented due to discrimination. Adjustment via affirmative action is likely to be helpful in this case. On the other hand, Asians are discriminated against when they appear to be high performing, which means they are unlikely to be at the margin. So, Asians who are at the margin of college admission may not be facing discrimination. Asians would need another policy to be protected from the types of discrimination they face.

The rest of the paper consists as follows: We review the existing literature. Next, we propose terminologies and an experimental design for workers and managers. Next, we describe the empirical strategy. Then, we show the results for workers and managers. And we separately demonstrate the ICT study. We conclude with further discussion.

2. Literature Review

Even though the existing literature has many papers that study race discrimination, almost all of them compare only two groups.[2] Table 1 presents the number of race

[2]For example, Agan and Starr (2017), Arnold et al. (2018), Anwar et al. (2012), Altonji and Pierret (2001), and Charles and Guryan (2011) compare White and Black. Burgess and Greaves (2013), Alesina and La Ferrara (2014), and Goncalves and Mello (2021) compare Whites and non-Whites. Åslund et al. (2014) and Arai and Thoursie (2009) compare immigrants and native Swedish. Fershtman and Gneezy (2001) compares two groups in Israel. List (2004) and List (2006) also majorly focus on Whites and non-Whites but also consider other attributes simultaneously, such as age and gender. Haaland and Roth (2023) focuses on White and Black, showing that pro-black policy preference barely changes using an experiment. Szymanski (2000), Ondrich et al. (2003), and Wozniak (2015) compare Black and non-Black people.

Journal	Total #	Experiments		Field Studies	
		Binary	3+ groups	Binary	3+ groups
AER	13	4	1?	5	3
QJE	21	7	0	11	3
ECMA	2	1	0	1	0
ReStud	2	0	0	2	0
JPE	11	0	0	9	2
ReStat	24	4	0	17	3
JLE	10	0	0	10	0
AEJ:Applied	4	2	0	2	0
AEJ:Policy	1	0	0	1	0
	88	18	1	58	11

TABLE 1. Literature

discrimination publications in top journals from 1990 to 2022. There are only 12 publications out of 88 that consider more than two groups, and in experimental work, there was only one publication.

Also, even when multiple minorities are included, they are usually considered as one consolidated minority group rather than multiple separate groups.[3]. This approach of one minority could miss some important facets of discrimination, as stated in the previous section.

Still, there is a handful of research on racial discrimination with multiple minorities that did not treat all minorities as one group. Eyting (2023) and Shi and Zhu (2023) focus on the effect of one minority on the others. In contrast, we compare the majority as well. Rather than how minorities affect each other, we test how each race is treated independently. Similar to our view, Chan (2023) studies three race groups without focusing on the interactions between minorities. The differences with our study are that Chan (2023) mainly focuses on underpayment compared to Whites and the role of signals without eliciting beliefs directly. Another paper that considers three groups is Aaronson

[3]For example, Feigenberg and Miller (2022), Avenancio-León and Howard (2022), Antonovics and Knight (2009), and Anwar and Fang (2006) consider White, Black, and Hispanic groups. Holzer and Ihlanfeldt (1998) also consider the same three groups, but the main focus of the minority is Black people. Goldsmith et al. (2006) compares White and Black groups, but the Black group is divided into several subgroups based on their skin tone. Kreisman and Rangel (2015) also studies Black people's skin tone and resulting discrimination. Christensen and Timmins (2022) compares Whites, African Americans, Hispanics, and Asians. Fryer (2019) compares Whites and Black people and then compares Whites and Hispanics separately.

and Phelan (2022), studying job losses of Whites, Asians, and non-Asians of color. They show that the third group experienced more job loss. This also has a common ground with our work as it shows different outcomes for two minorities.

One important stream of literature in discrimination consists of resume audit studies (for example, Bertrand and Mullainathan (2004), Kline and Walters (2021), Kline et al. (2022)). Resume audit studies in the fields are useful, but some limitations could exist. First, resume audit studies could have a deception problem, since the resumes used are not real. Kessler et al. (2019) suggests a novel way to avoid the deception problem in an audit study, but their approach can raise new incentive problems. In our work, we use incentive-compatible payment methods and also use real resumes to avoid deception problems. In addition, researchers in field audit studies can observe only the call-back rate but not continuous ratings. In our experiment, we ask willingness to pay ranging from 0 to 20, so we have observations that can show discrimination at a finer level. Moreover, researchers in audit studies cannot observe beliefs, and thus, it is hard to disentangle statistical and taste-based discrimination. In this paper, we elicit beliefs directly so we can identify taste-based discrimination more than the existing audit study.

There are plenty of studies that focus on other aspects of discrimination. One of the most actively studied areas of discrimination is gender [4]. Even if it studies mostly binary cases by its nature, there are still some similarities. For example, we borrow the design for control treatment that does not have racial information, as in Exley and Nielsen (2023). Also, there are works that consider abstract groups rather than real identities. Our main focus is to show the existence of discrimination, but the abstract groups on discrimination with in-group and out-group biases could have implications for managers' behaviors. For example, Chen and Li (2009) shows in-group and out-group biases using abstract groups. They found that people prefer agents from the same group. Similarly,

[4]For example, Babcock et al. (2017), Niederle and Vesterlund (2007), Exley and Kessler (2022), Reuben et al. (2014) and other works.

Heap and Zizzo (2009) shows discrimination against outsiders of their own group, resulting in welfare loss using abstract groups. One of our results shows a difference in beliefs that favors low-performing groups but not specific racial groups. Some work on abstract groups that study beliefs does not impose any intrinsic preference on managers regarding workers' group identity other than performance. Thus, such works can be related to our results. For instance, Mengel and Campos Mercade (2022) shows that employers who neglect signals discriminate more against the disadvantaged group even when group identities do not include any demographic information. Likewise, Esponda et al. (2023) shows that people update to a more extreme extent than Bayesian posterior when two groups are presented together, even when group identities are abstract.

There are several theoretical works for explaining discrimination. Those works can suggest the mechanism behind the discrimination behaviors of managers. Fryer (2007) shows how statistical discrimination works in the dynamic game and whether a discriminated-against player can overcome it. Bohren et al. (2019b) also studies the dynamic setting and suggests evidence from field experiments. Their results are similar in that the initially discriminated against group could reverse it in the dynamic setting. Bartoš et al. (2016) suggests a model of information acquisition to explain discrimination and provides evidence from the field experiment. Methodologically, Arnold et al. (2022) develops new quasi-experimental tools to estimate the disparity between Whites and Black people in bail decisions. They also find evidence of both racial bias and statistical discrimination. For theoretical works that are indirectly related to discrimination, Frankel (2021) shows managers' decisions on hiring as a function of test scores on resumes in the labor market with a principal-agent model. Akerlof and Kranton (2000) modeled how one's group identity can affect economic behaviors and outcomes, which has an implication in the labor or education market, such as exclusion. Another related paper is Bordalo et al. (2016). They suggested a model of stereotype and showed that stereotypes could be distorted depending on a reference group and context-dependent. They also show experiment results both in an abstract setting and a political setting.

3. Worker-Manager Experimental Design and Terminology

The experiment comprises three phases. We refer to participants in each phase as 'workers' for Phase 1 and 'managers' for Phase 2. The primary objective of Phase 1 is to create a pool of workers who will be assessed or hired by the managers in Phase 2. Phase 2 is specifically designed to measure both beliefs and willingness to pay wages of managers. Moreover, to investigate whether such discriminatory patterns are related to race, we conducted two between-subject treatments: a Race-Revealed treatment and a Race-Blind control. Finally, in a separate experiment, we examine whether individuals are willing to address and correct discrimination in the workplace. The design and results of this experiment are provided in Section 6. Screenshots of both experiments are provided in Appendix A.

3.1. *Terminology*

We provide definitions from the existing literature here, with slight modifications as needed.

Definition 1 (Statistical Discrimination)**.** Let a worker be a High type if they pass a threshold score for a task. Let $P_g(H)$ be a probability belief that a random member of group g is a High type. For any groups g_1 and g_2, if $P_{g_1}(H) \neq P_{g_2}(H)$, then we say statistical discrimination exists.

This definition of statistical discrimination does not imply any accuracy or inaccuracy of beliefs. Arrow (1974) and Phelps (1972) define statistical discrimination as a prediction of the quality of workers. Here, that problem is simply $P_g(H)$. Importantly, statistical discrimination is a property of beliefs only, though, of course, statistical discrimination can cause differences in wages or employment success. We refer to taste-based discrimination as a situation where wages differ even if beliefs do not.

Definition 2 (Taste-Based Discrimination)**.** Let W_g be the offered wage for a representative of group g. For two groups, g_1 and g_2, we say taste-based discrimination exists if $P_{g_1}(H) = P_{g_2}(H)$ but $W_{g_1} \neq W_{g_2}$.

Becker (1971) defined taste-based discrimination as having an implicit cost for hiring a certain group. We take that definition, especially for the groups that are believed to have the same quality; the group identities alone lead to different wages.

Note that discrimination can act in both ways. One group can be paid less due to both taste-baste and statistical discrimination compared to others. For example, even after fixing taste-based discrimination, one group can still be paid less than the others due to generally low beliefs about them. By measuring both beliefs and wages, our experiment can measure both types of discrimination.

3.2. *Phase 1: The Worker Study*

Workers are asked to provide information that will be used to create their resumes. The information includes their hometown, high school activities, and simple computer skills. We restrict workers to those who have reported their race and gender information previously. Subsequently, they are assigned three tasks, each designed to measure different skills. The three tasks assess their math skills, social skills, and a combination of both skills. We refer to the last one as 'combined skills' since they require proficiency in both math and social skills.

The first task measures math skills. Workers solve ten multiple-choice questions. For each question, they have 15 seconds to solve. For each correct answer, they get one point.[5]

For the social skills measurement, workers play a coordination game in which they assess the social appropriateness of different behaviors. The design is adopted from Krupka and Weber (2013). They are given a scenario of the dictator game and assess

[5] We modified questions in Bohren et al. (2019a) and the ASVAB test, following Exley and Nielsen (2023).

the social appropriateness of each of the dictator's strategies using a four-point scale.[6] For each strategy, they earn a point if their assessment matches the mode response of all workers. While Krupka and Weber's primary focus is on measuring social norms, this design also allows us to capture their ability to predict others' perceptions of norms. Hence, we interpret this game as a measure of social skills. The social task measures just one particular type of social skill. Of course, we could use other tasks that would measure other types of social skills, but for our purposes, what matters is simply that we have different tasks measuring different types of skills.

In the combined task, workers play a continuous time centipede game in which they are presented with a formula for calculating the payoffs in each second, rather than the explicit payoff amount. To maximize their payoff, they should be able to calculate payoffs for each second from these rules, which requires math skills. They play the game against all the other workers. Instead of playing the game individually, they choose only one stopping point, and this point is implemented in all of the games. The average earnings from all these games then determine their final payoff. Deciding on a stopping point demands social skills, as workers are required to predict others' average behaviors to optimize their own strategy. As in the social task, we just focus on one type of task that requires both math and social skills, as what matters is having different tasks for different types of skills. The exact task used is not particularly important for our study.

For each task, workers are classified as a High type if they get a score greater than or equal to some threshold score and a Low type otherwise. Thus, types are defined based on absolute performance, not relative. The details of the tasks and the threshold values are in Appendix A.

For this 10-minute study the workers receive a completion fee of $1.50, and they also have the opportunity to earn a bonus payment. The bonus payment is determined by their performance from the three tasks. One of the three tasks is chosen for the bonus payment, which can be up to $1.

[6]very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, very socially appropriate

3.3. *Phase 2: The Manager Study*

The manager study consists of two treatments: (1) the Race-Revealed treatment and (2) the Race-Blind control.

In the Race-Revealed treatment, each manager has 14 resumes to evaluate. Before evaluation, they are informed about the tasks workers completed. For the first six resumes, high school activities are not shown. We elicit the manager's belief that each is a High type worker for each of the tasks. We refer to this as a prior belief since high school activities are not shown, and high school activities serve as a signal of the worker's abilities. Managers evaluate resumes from each race. The subsequent four resumes provide high school activities. For these, we elicit the manager's posterior beliefs about each worker being a High type in each task. The details of the signals and how they work will be described in subsequent paragraphs. Finally, for the last four resumes we elicit the wages managers are willing to pay for each of three tasks. One of the 14 resumes is then randomly chosen for payment.

The first six resumes include race, gender, and the distractor information. Distractor information is hometown, MS Word proficiency, and Adobe Illustrator proficiency. None of these distractor elements are relevant to the three tasks and demographic traits.[7] Each of the six resumes has distinct demographic information, ensuring a non-repetitive representation. In other words, the resumes have all combinations from $\{White, Black, Asian\} \times \{Female, Male\}$, and thus each resume has a unique combination of race and gender. Thus, the beliefs we elicit in this stage give a prior probability of being High type for each race with minimal information.

To avoid confusion, we clarify what we mean by prior and posterior. To distinguish beliefs elicited in these first six resumes and the following four resumes with additional information, we call the beliefs from the first six resumes the "prior". Note that when

[7]One might think that MS Word and Adobe Illustrator proficiency could be perceived to be related to the three skills. However, the distributions across skill levels were similar across races. Also, the control treatment further eliminates those concerns as we compare only within one race; Race-Revealed and Race Control have essentially the same contents within each race. Details will be described in the following paragraphs.

it comes to the "prior", it is a belief based only on demographic traits. That means we do not refer to as race (and gender) information as a signal and use the terminology *prior* to describe the probability of a certain race being of High type without further signal information. Posteriors are then the beliefs elicitated from resumes that reveal the worker's high school activities, which may be informatibe about their skills.

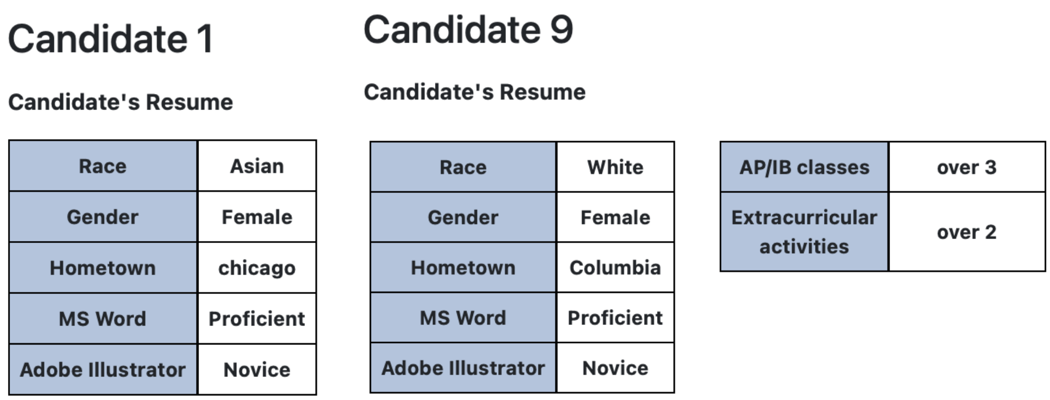
The subsequent four resumes for which we elicit the posterior contain two additional items: (1) the number of AP/IB classes the individuals took in high school, and (2) the number of extracurricular activities they were involved in during high school. The former is potentially associated with math skills, while the latter may be potentially linked to social skills, and thus both could be related to the combined skills task. Both items are presented in two ranges.[8] We regard these two items as signals. This allows us to interpret them as either a high signal (h) or a low signal (l), leading to four distinct signal combinations: $\{hh, hl, lh, ll\}$. Each manager is provided with one resume per signal combination. Furthermore, the demographic identities of the workers are mutually exclusive, ensuring that each *race*×*gender* combination is unique and not replicated among the four resumes. Orders of the four resumes are randomized. Both prior and posterior elicitation are incentivized by the binarized quadratic scoring rule. Their payments only depend on their own beliefs, and those reports on beliefs do *not* affect the workers. Thus, belief elicitations are not affected by taste-based discrimination.

The last four resumes follow the same format as the posterior resumes, but are used to elicit the wages managers are willing to pay. Beliefs are not elicited for these last four resumes. Again, managers see on resume for each signal combination, though they are different from the previous resumes. For example, for the hh signal resume, a manager could have the $White \times Male$ for posterior resume while having $Asian \times Female$ for the willingness to pay resume. For each resume, managers report their willingness to pay for all three tasks. Willingness to pay elicitation is incentivized by the Becker–DeGroot–Marschak method. Specifically, if a worker is a High type, then

[8]For the number of AP/IB classes, 0-3 or over 3. For the number of extracurricular activities, 0-2 or over 2.

# of resumes	Elicited Object	Information
6	Prior	Race+Gender+Distractor
4	Posterior	Race+Gender+Distractor + AP/IB classes + Ext. Curr. Activities
4	Willingness to Pay	Race+Gender+Distractor + AP/IB classes (signal 1) + Ext. Curr. Activities (signal 2)

TABLE 2. Race-Revealed Treatment Summary



(A) Resume for prior (B) Resume for posterior/willingness to pay

FIGURE 1. Resume examples

the manager earns a revenue of 20 from hiring this worker. If the worker is a Low type, then they only earn a revenue of 10. The manager chooses their willingness to pay, and then a random number (which represent the actual wage) is drawn from the range of 0 to 20. If the willingness to pay is greater than or equal to the drawn number, the manager gets *revenue−the actual wage*, and the worker receives the actual wage. If the willingness to pay is less than the actual wage, both are paid 0. One notable thing here is that we actually pay the wage to the worker if a resume from the willingness to pay task is chosen for payment. Managers are also informed of this information. Thus, this task can be affected by taste-based discrimination.

The Race-Revealed treatment allow us to measure how belief-updating behaviors and wage choices are different for different race groups. However, a valid concern arises regarding whether divergent reactions across racial groups are solely attributable to race, or whether they could be due to other factors that happen to differ across races.

For example, consider a scenario where the prior belief of Asians of a High type in the math task is 65%, while it is 60% for Black people. Now, suppose that managers update their posterior to 90% for Asians and 65% for Black people upon receiving an h signal. In this instance, it could be that the differing posteriors are due to statistical discrimination but it could also just be that slightly different priors lead to very different posteriors regardless of race. Similarly, consider a situation where posterior beliefs for Asians and Whites are 65% and 60%, respectively, but the willingness to pay for both groups is the same. Since Asians receive the same amount as Whites even if they are perceived as more likely to be a High type, it might seem like evidence of taste-based discrimination. However, it also could be that the function that maps belief to willingness to pay is flat within the range of 60% to 65%. In addition, distractor information can also cause some systematic biases that we cannot observe directly.

These concerns bring up the necessity for the control, which omits racial information from resumes. Managers in the Race-Blind control have the same pool of workers and have the same tasks. The only difference between the two treatments is the absence of racial information on the resumes in the control. If updating patterns and wage choices are the same between the Race-Revealed treatment and the Race-Blind treatment, then we can rule out race as a cause of those patterns. This design for control treatment is similar to Exley and Nielsen (2023). In their case, they replaced gender groups with abstract group name control.

Phase 2 lasted 15 minutes on average. The completion fee is $4; depending on their choices, subjects can earn an additional bonus payment of up to $2.

3.4. *Procedures*

We recruited 300 workers, 289 Race-Revealed treatment managers, and 199 Race-Blind control managers. All subjects were recruited from Prolific. We dropped 18 workers who got 0s for the multiple choice math task or chose 0 seconds as the stopping point in the combined task (which generate a payment of 0) since we regard those as indicating intentionally low effort.

4. Empirical Strategies

Before we proceed with the results, we demonstrate our empirical strategies. The strategies are used to clearly identify taste-based discrimination As we mentioned in section 3.1, measurement of taste-based racial discrimination requires the manager to have the same beliefs between two groups. However, we do observe different beliefs across races. But we can control this by using the Race-Blind that do generate a “race-free" wage that managers would pay for each possible belief. We can then compare the Race-Revealed manager’s actual wage to the “race-free" wage they would have paid given their stated beliefs. The difference is a measure of taste-based discrimination. Specifically, we estimate this “race-free" wage as a mapping from beliefs to wages. And we estimate this for each signal combination for each race. In this section, we describe the procedures for estimating such mappings.

The estimation requires two steps: 1. matching data, and 2. non-parametric censored regression. Using simulations, we then validate this method. The simulation steps and results are presented in Appendix B.

First, recall that managers saw different resumes when providing posterior beliefs and when providing their willingness to pay (or, wage). This means we cannot directly observe their willingness to pay for each observed belief. Instead, we need to match managers’ beliefs and wages at the population level, rather than the individual level. To do this, we take all resumes with the same signal (for example, hh) and match the lowest posterior with the lowest wage, the second-lowest posterior with the second-lowest wage, and so on.[9] We do this for each of the three tasks separately.

[9]If the number of observations of the posterior and wage differ we randomly drop observations from the larger one to equalize the two samples.

Next, we take each of these matched data sets and estimate the following equation by grid-searching for the coefficients that minimize the mean squared error of the matched data:

$$w_i = \min\{\beta_0 + \beta_1 \times post_i + \epsilon_i, 20\}$$

The upper bound of 20 represents the maximum revenue managers could earn. This method is a non-parametric version of the Tobit regression. [10]

Finally, we used our simulation to de-bias these estimates. First, we estimated coefficients with the simulated data. For each estimated coefficient we know its true value, so we can observe the mapping from true values to estimated values. Then, with the coefficients from the real data, we back out the closest estimates from the simulation that minimize percentage deviations.

From this we can generate a "race-free" wage each manager in the Race-Revealed treatment would have paid given their beliefs.

5. Worker-Manager Results

This section provides the results of Phase 1 and Phase 2. Unless otherwise stated, we employ the permutation test for the analysis.

Our main results are Result 5 and Result 6, presented in section 5.2.2.

5.1. *Phase 1: Worker Results*

This subsection gives brief overviews of Phase 1 workers' performance. The next subsection discusses the workers' performance in more detail along with managers' evaluations.

[10] The estimated coefficients are close to Tobit model coefficients. Also, McDonald and Moffitt (1980) shows that the Tobit-regression coefficients are decomposed into two parts. Rather than back out the part that is of interest from Tobit coefficients, we directly estimate our main interest model coefficients.

Table 3 summarizes the workers' results. The top row contains the workers' actual frequency of being the High type, broken down by race. We use the Pearson's Chi-Squared test for statistical tests since the type is a binary classification. The second row reports the Race-Revealed managers' average prior beliefs for comparison. Each cell also provides the ordering of the three races in terms of either actual performance or prior beliefs.

	Math Task			Social Task			Combined Task		
	White	Black	Asian	White	Black	Asian	White	Black	Asian
Worker (Actual)	15.79 %	9.78%	38.95%	56.84%	42.39%	53.68%	45.26%	34.78%	45.32%
		A> W~B			A~ W>B			A~ W>B	
Manager (Avg. Belief)	52.62%	52.26%	67.10%	57.42%	57.42%	54.16%	56.96%	57.99%	63.35%
		A> W~B			W~B>A			A> W~B	

TABLE 3. % of workers that are the High type and Race-Revealed managers' prior
[1] W=White, B=Black, A=Asian
[2] >: differences are statistically significant, ~: differences are not statistically significant

Managers over-estimate workers' performance in general. In the orderings, managers have wrong beliefs in the social and combined tasks. Specifically, Asians are regarded as less likely to be a High type in the social task compared to Whites and Black people. However, Whites and Asians' actual performances are not significantly different. For the combined task, Asians and Whites perform the same, but managers' beliefs are higher for Asians.

Result 1. Managers overestimate workers' performance. Furthermore, their ordering of the performance of races is incorrect, favoring Black people over Asians on the social task and underestimating Whites in the combined task.

5.2. *Phase 2: Manager Results*

In this section, we explore our main variables of interest: Managers' beliefs and wages. In these analyses, we do not control the race identity of managers since Asians in the

Prolific pool have substantially higher education levels. This might reflect the real education level of the entire population but also create a confound since differences in beliefs or wages could be due to education differences or other unobserved factors.[11]

5.2.1. *Beliefs*

In this subsection, we will show that managers do find race information to be informative for Asians, but generally not for Whites or Black people. In terms of absolute levels, we find that beliefs on the math and combined tasks are consistently much higher than actual performance for all three races. Finally, although Asians perform substantially higher on the math and combined tasks, managers' beliefs underestimate the magnitude of this difference.

To see these results, consider the Math task results shown in Figure 2. Bars are grouped by race, with the left bar showing the actual percentage of workers of that race who were the High type, the middle bar showing the average beliefs of managers in the Race-Blind treatment, and the right bar showing the average beliefs of managers in the Race-Revealed treatment. The three panels show the prior belief, the posterior belief after an h signal, and the posterior belief after an l signal. Asterisks on the bars indicate statistical significance between the Race-Blind and Race-Revealed beliefs. This figure shows that Race-Revealed managers have significantly higher beliefs for the Math task when it is revealed that a worker is Asian, but not when they are White or Black. Thus, managers view race as an informative signal only for Asians. This is also true for the social task (Figure 3)—though now being Asian is seen as a negative signal while being White is positive—and for the combined task (Figure 4), where again managers update positively for Asians.

Next, we compare managers' beliefs to workers' actual performance. On the Math task, the beliefs of managers are uniformly higher than actual performance for all three

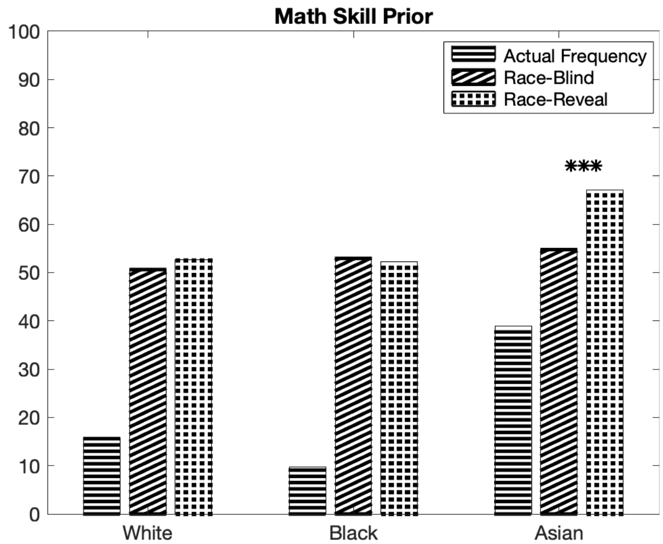
[11]Though, we conducted a basic analysis when controlling the managers' racial identity. The power of the test is low, but we found that there is no significant difference in results due to the manager's race. Detailed results are in the Appendix.

races, with all p-values below 0.01. Although Asian workers have substantially higher performance on this task, managers' beliefs underestimate this difference. Thus, although managers have the highest overall beliefs for Asians, the gap between beliefs and performance is actually the lowest. In other words, the lower-performing groups benefit more from the miscalibrated beliefs of managers and from the fact that managers don't update much when race is revealed.

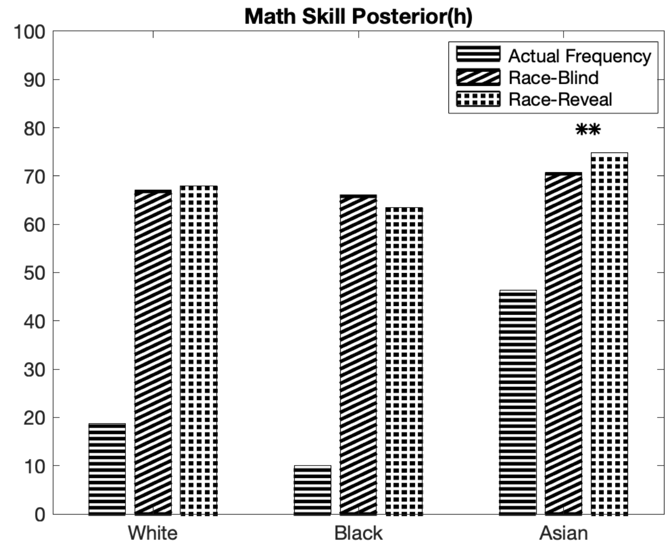
A similar pattern exists for the social task case. As Figure 3 illustrates, Asians and Whites are the high-performing groups on this task.[12] Statistically significant differences between Race-Blind and Race-Revealed beliefs exist only in the White and Asian priors, with a positive update for Whites and a negative update for Asians. In this case, those differences make managers' beliefs even closer to workers' actual performance. For all beliefs (priors and posteriors), Black people get the most benefits as the differences between the Race-Revealed beliefs and actual performance are the largest. However, again, this is due to miscalibrated high beliefs that exist in Race-Blind beliefs as well. For the posterior with an h signal, Asians are the best-performing group but get the least benefit. Similarly, Whites get the least benefit in the posterior with an l signal while they are the best-performing group in this case. All these observations are consistent with the math results, where a high-performing group gets the least benefit, and low-performing groups get the most benefit due to general miscalibration that is prevalent in Race-Blind beliefs as well.

These same patterns also appear in the combined task (Figure 4), with the exception of prior beliefs shown in Panel 4a. The difference between Asian and White workers' actual performance is not statistically significant. However, the difference in the Race-Revealed prior and actual performance is larger for Asians (on average, the differences

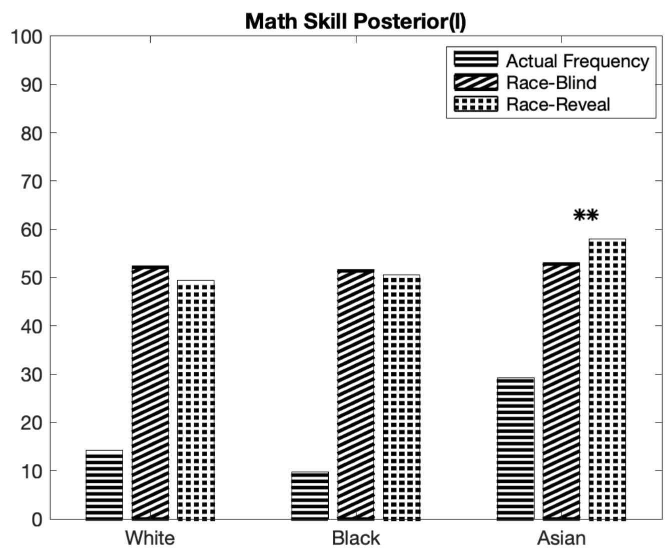
[12] No statistically significant difference for the overall population (depicted in Prior). In the social task case, an h signal means 'over 2' Extra-curricular activities, and a l signal means '0-2' Extra-curricular activities. Asians are higher with h-signal posterior, and Whites are higher with l-signal posterior.



(A) Math task prior



(B) Math task posterior after h signal



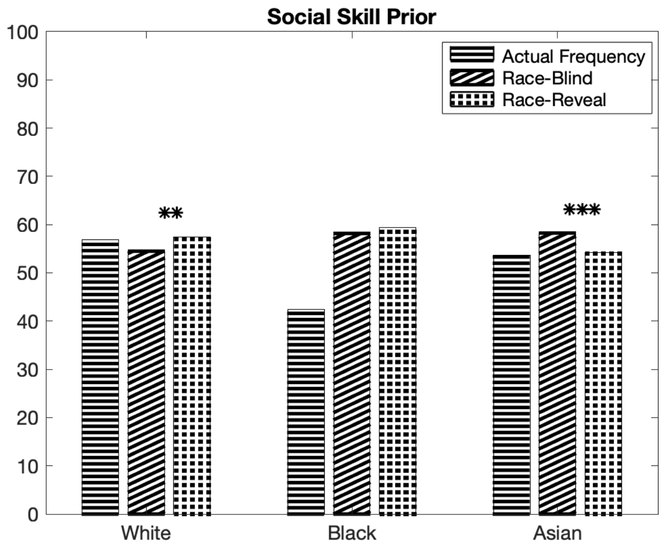
(C) Math task posterior after l signal

[1] Asterisks indicate statistical differences between Race-Reveal and Race-Blind.
[2] * p<0.1, ** p<0.05, *** p<0.01.

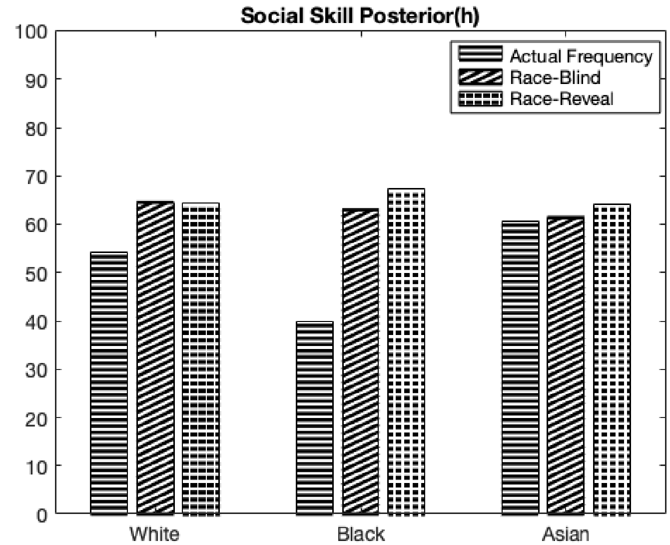
FIGURE 2. Math Beliefs

for Whites and Asians are 11.70 and 17.03, respectively, and these differences are different with a p-value of 0.000).[13] Still, the lowest-performing group gets the biggest

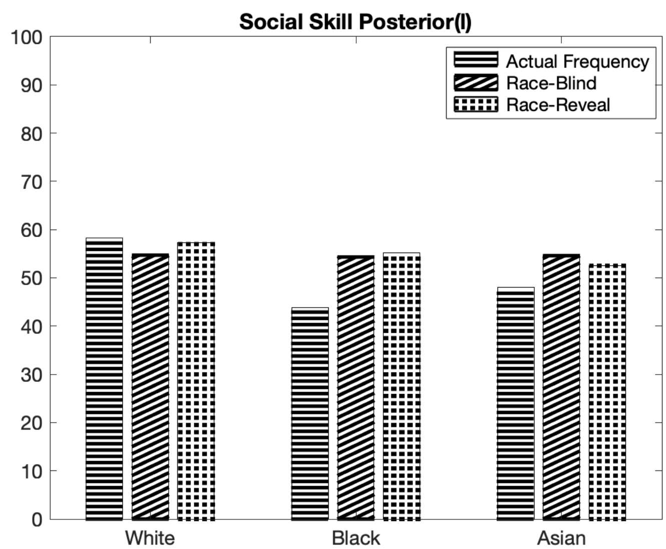
[13]However, this result is still partially explained by the same reason: high beliefs in Race-Blind. The mean prior in Race-Blind is 55.67% for Whites and 58.12% for Asians. The difference is statistically significant at the margin (p-value=0.09).



(A) Social task prior



(B) Social task posterior after h signal



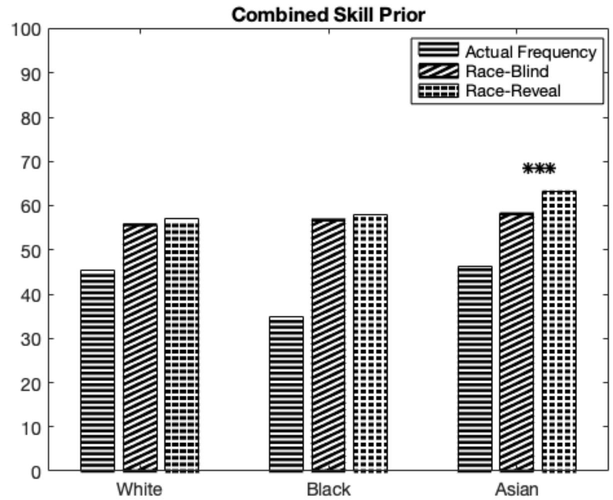
(C) Social task posterior after l signal

[1] Asterisks indicate statistical differences between Race-Reveal and Race-Blind.
[2] * p<0.1, ** p<0.05, *** p<0.01.

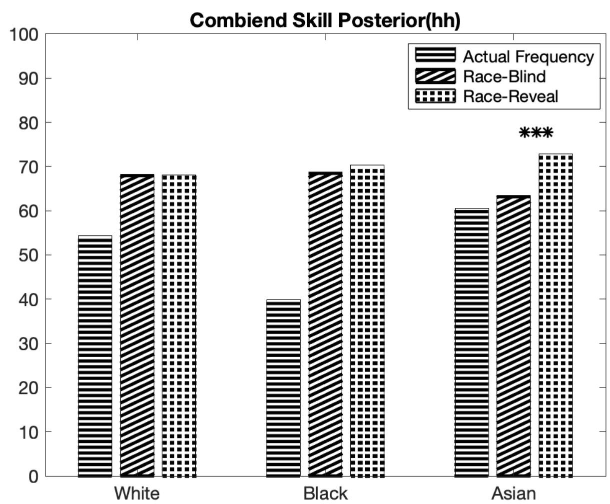
FIGURE 3. Social task beliefs

benefits, which is largely consistent with the math and social task results. And we find this same pattern in posterior beliefs, as shown in Panels 4b to 4e.

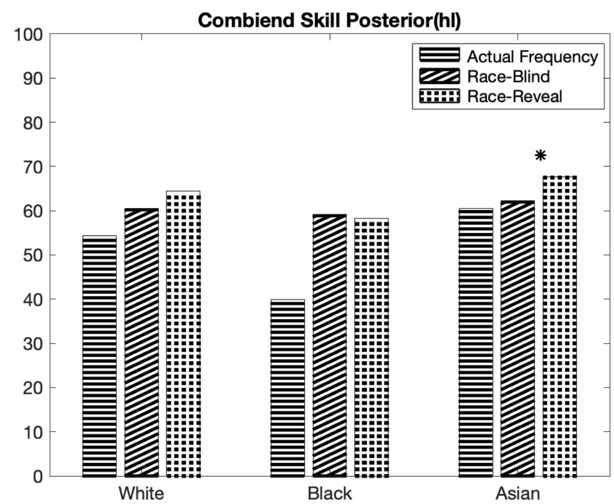
Result 2. Managers generally only view race as informative for Asians.



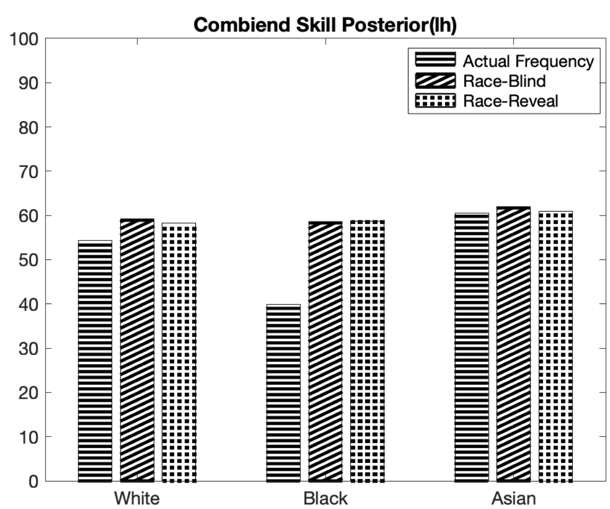
(A) Combined Skill Prior



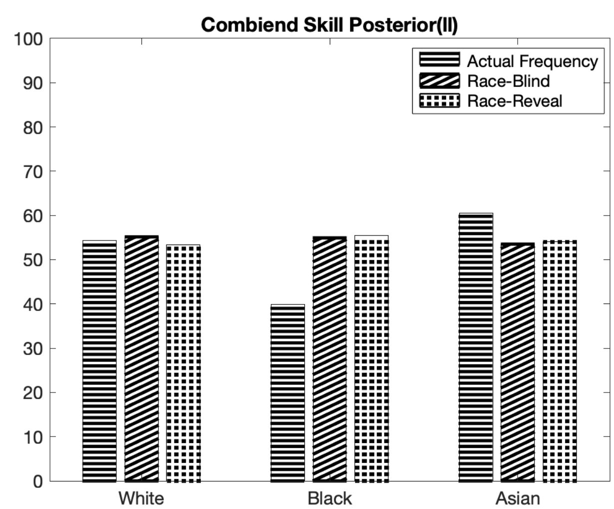
(B) Combined Skill Posterior hh



(C) Combined Skill Posterior hl



(D) Combined Skill Posterior lh



(E) Combined Skill Posterior ll

[1] Asterisks indicate statistical differences between Race-Revealed and Race-Blind.
[2] * p<0.1, ** p<0.05, *** p<0.01.

FIGURE 4. Combined Skill Beliefs

Result 3. Managers' race-blind beliefs are generally too high and do not update downward when race is revealed for low-performing groups.

	White	Black	Asian
Math Skill h signal	68.00%	63.53%	74.79%
		A>***W>**B	
Math Skill l signal	49.41%	50.57%	57.98%
		A>***W~B	
Social Skill h signal	64.38%	67.17%	64.10%
		W~B~A	
Social Skill l signal	57.30%	55.21%	52.57%
		W~B>**A	
Combined Skill hh signal	68.06%	70.38%	72.81%
		W~B~A	
Combined Skill hl signal	64.42%	58.24%	67.60%
		A~W>*** B	
Combined Skill lh signal	58.26%	58.65%	60.89%
		W~B~A	
Combined Skill ll signal	53.43%	55.46%	54.03%
		W~B~A	

[1] W=White, B=Black, A=Asian
[2] >: differences are statistically significant, ~: differences are not statistically significant
[3] * p<0.1, ** p<0.05, *** p<0.01.

TABLE 4. Race-Revealed managers' posteriors

Since we confirm there is no favor in belief formation due to race, we now compare beliefs from the Race-Revealed treatment to find evidence of statistical discrimination. Table 4 shows the existence of statistical discrimination (beliefs being different across races). Each number is the Race-Revealed managers' mean beliefs. The statistical significance of differences is indicated with an inequality sign. As the results suggest, statistical discrimination does not exist in half of the cases. In the other cases where statistical discrimination exists, there is no clear pattern.

Result 4. Statistical discrimination exists (beliefs differ across races) in only half of the cases. Beliefs are higher for Asians on the math task, which matches actual performance. Otherwise, there is no clear pattern.

5.2.2. *Wages*

We compare the hypothetical Race-Blind wage that is estimated via the method described in Section 4 and the actual wage from the Race-Revealed treatment to measure taste-based discrimination. From now on, we call a hypothetical Race-Blind wage a Race-Blind wage for convenience.

Our empirical approach is valid for several reasons. First, beliefs are not exactly the same: (1) for the instances where there are statistically significant differences in beliefs between Race-Revealed and Race-Blind treatments, we cannot directly compare wages. (2) Even when the differences in beliefs are not statistically significant, it does not imply that they are exactly the same. Small differences in beliefs can still affect wage decisions. Second, we show that the beliefs are not biased due to racial preference. Thus, using Race-Revealed beliefs to calculate Race-Blind wages clearly identifies taste-based discrimination.

Henceforth, we say a group is “underpaid" if the hypothetical Race-Blind wage is higher than the actual wage from Race-Revealed managers, and “overpaid" in the opposite case. Thus, if one race is underpaid, that means the racial group faces taste-based discrimination. We do not directly compare wages across races. Distractor information could affect wage decisions or add noise differently across races. Thus, it is more rigorous to focus only on differences between Race-Revealed and Race-Blind wages within a race and compare these differences across races.

Our main result is that Asians are mostly discriminated against with h signals, and Black people are mostly discriminated against with l signals. There are two cases where Whites are underpaid, but the probability mass is small, and the degree is less severe.

Math task h signal

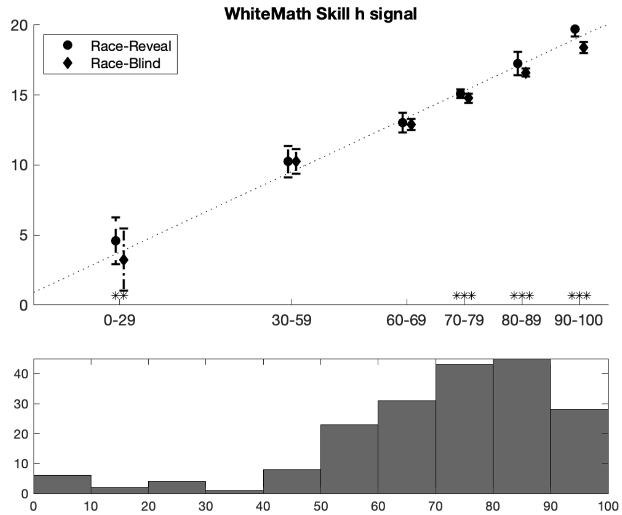
Figure 5 summarizes the result of the math task with an h signal. For each race, the upper panel depicts wage offers as a function of beliefs for both the Race-Blind and Race-Revealed treatment. The lower panel provides a histogram of Race-Revealed managers' posterior beliefs. For each indicated posterior range, Race-Revealed managers' wage offers shifted slightly left, and Race-Blind wages shifted slightly right for ease of comparison. We start with bins of posteriors set at intervals of 10 percentage points. If the number of observations within a bin is less than 10, we combine that range with a neighboring one.[14] We divide these bins in this way to compare races controlling for the same beliefs and to have adequate power in each bin. We applied the Holm–Bonferroni correction to avoid false-positives due to a large number of hypothesis tests.

As Figure 5a and Figure 5b depict, Whites and Black people are never underpaid for all posterior ranges, as the left point is never lower than the right point. However, Asians can be underpaid when they reveal their race. Specifically, Asians are, on average, underpaid in the 50-59% range by 13% and the 70-79% range by 3%, and 25% observations lie in these ranges.

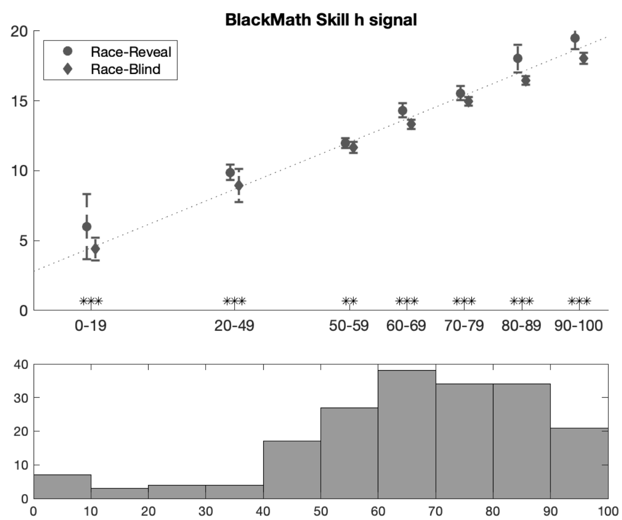
Asians do benefit in the 90-100% range, though the other two races also benefit in the same range. In this range, Whites, Black people, and Asians are overpaid by 7.16%, 8.03%, and 6.82%, respectively. Overall, 66.84% of Whites and 100% of Black people are in overpayment posterior ranges, while this number is only 22.16% for Asians.

To sum up, only Asians can be strictly underpaid when they reveal their race for the math task with an h signal. For Whites and Black people, both are not hurt by revealing their race, though the relative benefits are larger for Black people.

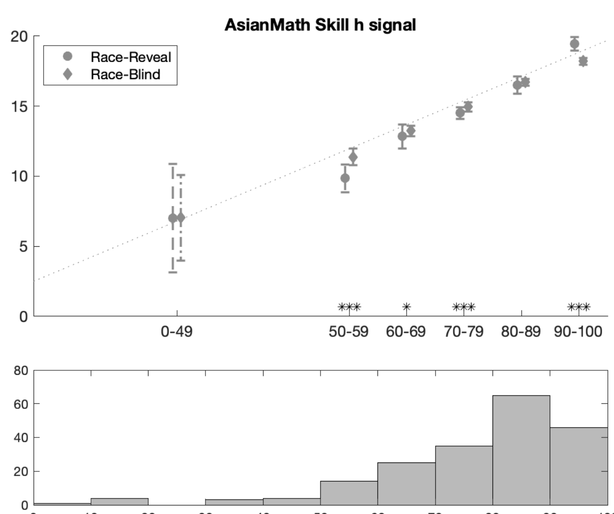
[14]We set this threshold of 10 observations for a reasonable statistical power. The permutation test suggests at least six samples per group with exhaustive permutations.



(A) White Math h signal



(B) Black Math h signal



(C) Asian Math h signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

FIGURE 5. Math task h signal

Math Task l Signal

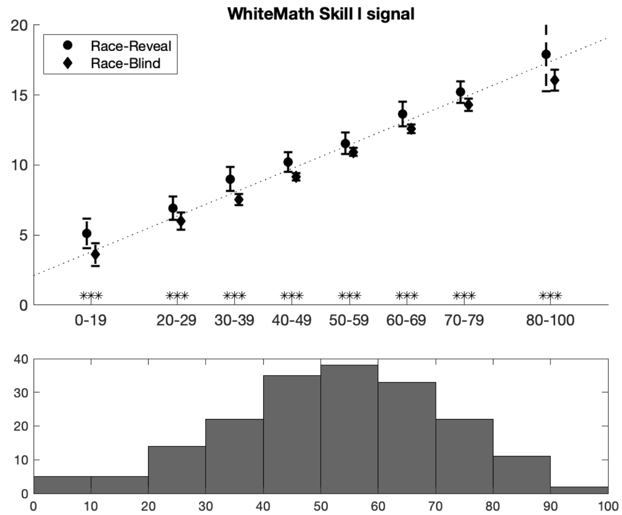
Figure 6 depicts the result of the math task with an l signal. Unlike the math task with an h signal, now Asians are not underpaid. They at least weakly benefit from revealing their race. Whites still benefit from revealing their race in all posterior ranges. On the other hand, Black people are underpaid by 1.75% in the 70-79% range, and 18.04% of observations are in this range.

For the 20-49% range, Black people benefit, and 36.08% of observations are in this range. Whites are overpaid by 14.72%, and Black people are overpaid by 11.57%.

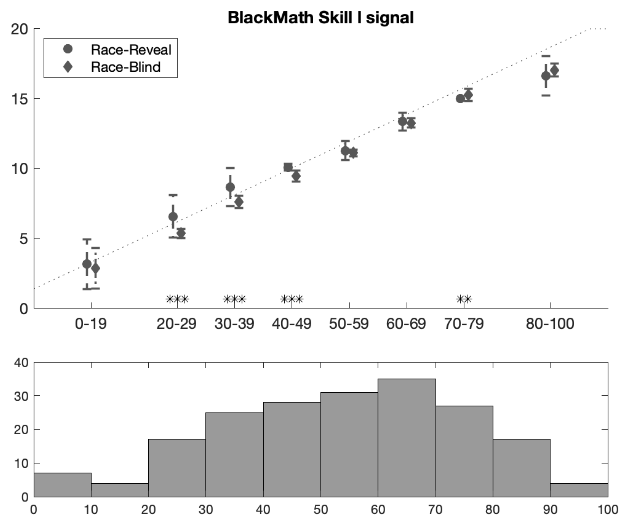
For the relative benefits between Asians and Whites, we first restrict our attention to the range 70-100% where both race groups get benefits. The percentages of Whites and Asians who are in this range are 18.72% and 32.64%, respectively. Whites are overpaid by 9.3% and Asians are overpaid by 7.36%. Thus, the relative benefits are not immediate in this range. However, as Whites benefit by revealing their race in all ranges, we can conclude that Whites are relatively more benefit than Asians.

To sum up, only Black people have the potential to be underpaid when they reveal their race with a l signal. Whites and Asians do not, and the relative benefit is larger for Whites.

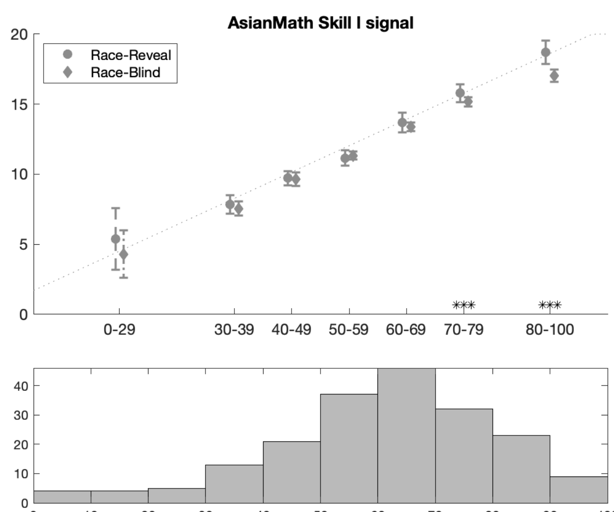
Result 5. In the math task, for certain ranges of beliefs, managers discriminate against Asians with an h signal and Blacks with a l signal. For some ranges of beliefs, Black people with an h signal benefit from having their race revealed. Finally, Whites are never discriminated against for any signal or any range of beliefs.



(A) White Math l signal



(B) Black Math l signal



(C) Asian Math l signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

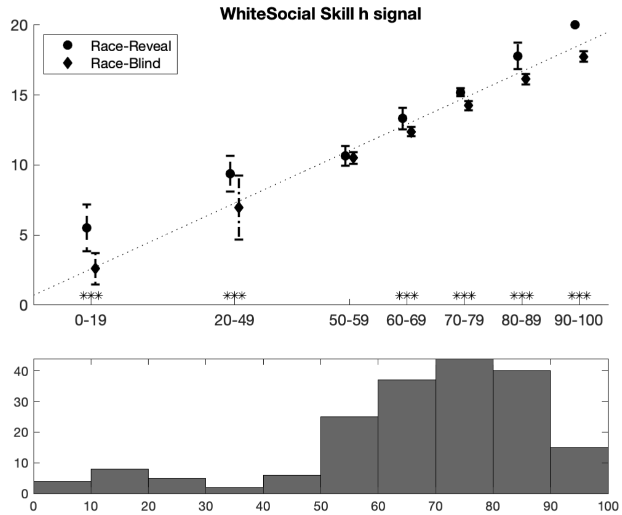
FIGURE 6. Math skill l signal

Social Task h Signal

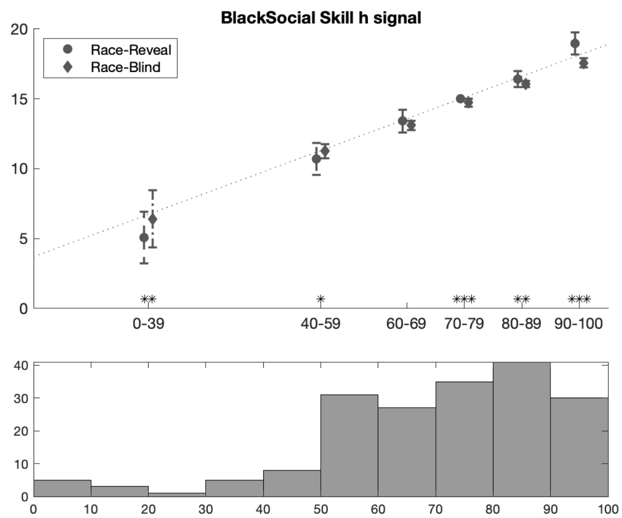
For the Social task with an h signal, a similar pattern to the Math skill is presented: Asians are underpaid the most.

As Figure 7a shows, Whites are never underpaid. Moreover, they are strictly overpaid other than in the 50-59% range. This is not the case for Black people and Asians. Unlike the Math skill case, now Black people also get underpaid in the 0-39% range (Figure 7b) by 20.93%. However, only 7.5% of the observations are in this range. For Asians, as depicted in Figure 7c, more ranges of beliefs have lower wages when their race is revealed. In the 40-69% range (specifically, 40-49%, 50-59%, and 60-69% ranges), Asians are underpaid, and 40.39% of observations are in this range. And they are underpaid by 6.86%. Thus, racial discrimination exists both against Black people and Asians with an h signal, but to a degree that is more severe for Asians.

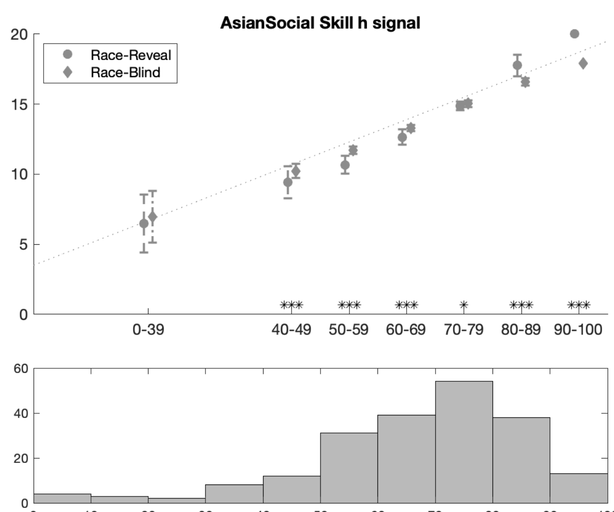
In the range of 80-100%, all three races are overpaid. In this range, Whites are overpaid by 11/64%, Black people are overpaid by 5.40%, and Asians are overpaid by 9.61%.



(A) White Social task h signal



(B) Black Social task h signal



(C) Asian Social task h signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

FIGURE 7. Social task h signal

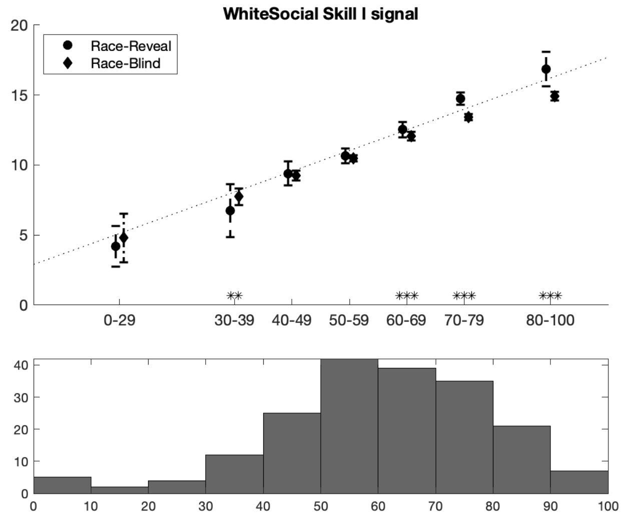
Social Task l Signal

Like the h signal case, the social task with an l signal also shows a similar result with the math task with an l signal: Black people are underpaid the most.

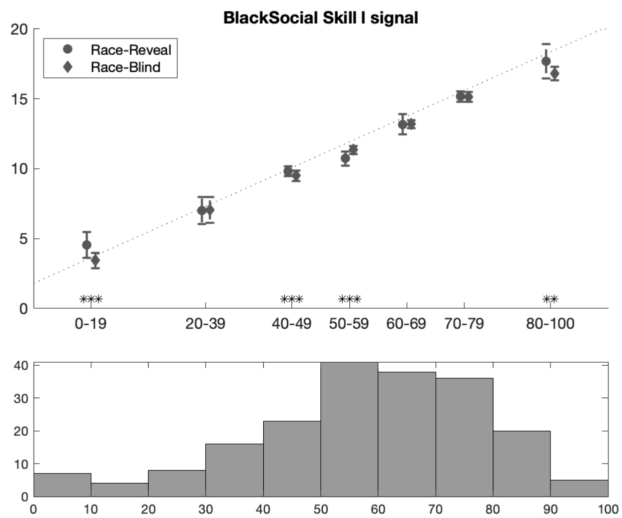
Asians are never underpaid for all ranges, as Figure 8c illustrates. Whites are underpaid in the 30-39% range (by 12.87%), and Black people are underpaid in the 50-59% range (by 5.41%). The percentage of observations that are in the underpayment range is 6.25% and 20.81% for Whites and Black people, respectively. It might not be immediate, but the result implies that overall underpayment is more severe for Black people.

Also, for the range 80-100% where all races are overpaid, Black people get the smallest relative benefits. Whites are overpaid by 12.91%, Asians are overpaid by 17.13%, while Black people are only overpaid by 5.27%.

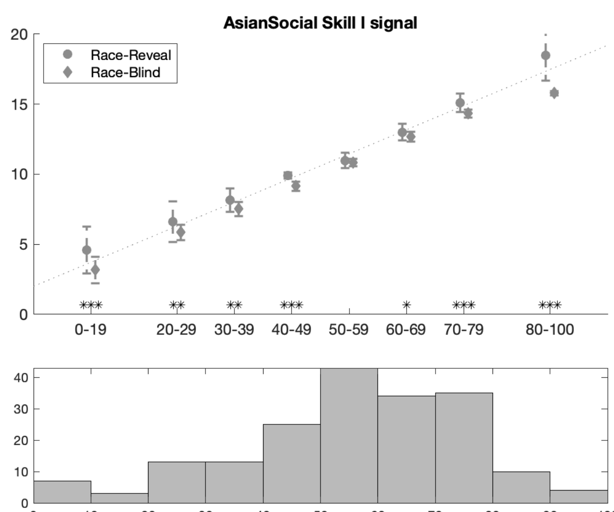
Result 6. Discrimination in the social task is similar to the math task: For certain ranges of beliefs, managers discriminate against Asians with an h signal and Black people with a l signal. Here, Asians with a l signal benefit from having their race revealed. Finally, Whites with a l signal are discriminated against for some ranges of beliefs, though the magnitude is relatively small.



(A) White social task l signal



(B) Black social task l signal



(C) Asian social task l signal

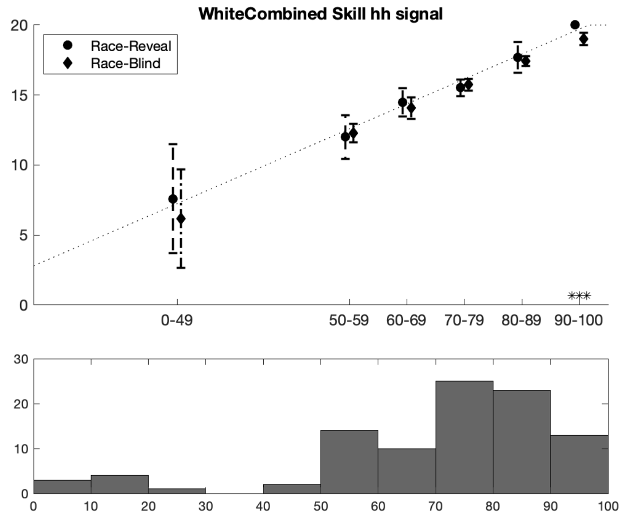
[1] * p<0.1, ** p<0.05, *** p<0.01.

FIGURE 8. Social task l signal

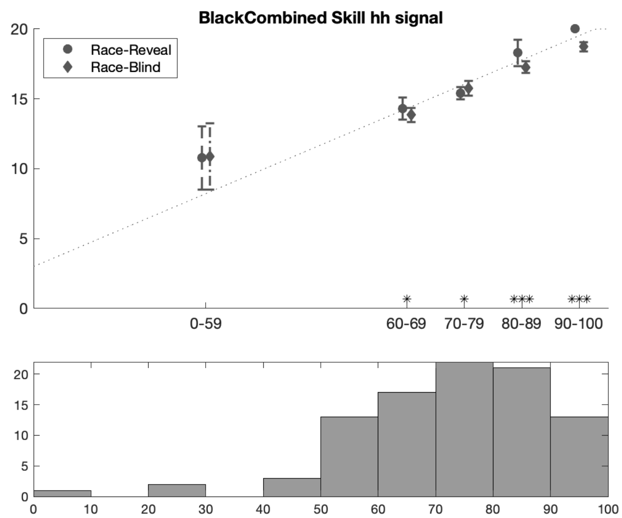
Combined Task

The combined task also follows a similar trend. For an hh signal, which is an unarguably high signal, Asians certainly get lower wages: 48.57% of observations are in the underpaid region, and they are underpaid by 11.23% on average. In contrast, Whites and Black people are always overpaid or at least get wages equal to the Race-Blind wage. For Whites, 23.75% of the observations are in the strictly overpaid range, and are overpaid by 5.19% on average. In the case of Black people, 51.42% of them are in the overpaid range, and are overpaid by 6.40% on average. The relative benefits are larger for Black people compared to Whites. These results are demonstrated in Figure 9.

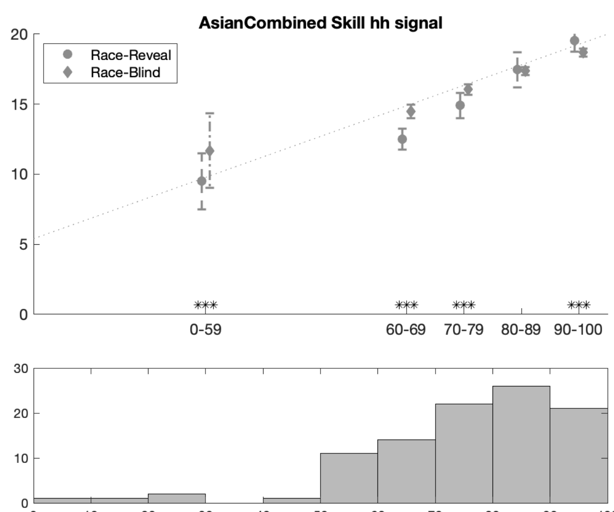
For a ll signal, which is obviously a low signal, all races are never underpaid. However, the relative benefits are the smallest for Black people. The results are demonstrated in Figure 10. For Whites, 75.93% are in the strictly overpaid range, with 7.64% overpayment on average. For Asians, the benefits of revealing race are even larger: 77.27% of observations are in the strictly overpaid range, with 11.62% average overpayment. For Black people, even if they are still overpaid, the degree is smaller. The strictly overpaid range contains 40% of observations, and they are overpaid by only 6.77%. Thus, these results conclude that Black people get the least benefit from revealing their race with a low signal.



(A) White combined task hh signal



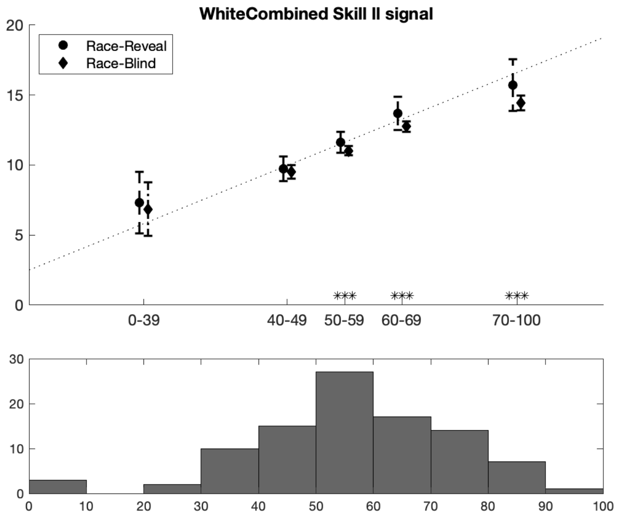
(B) Black combined task hh signal



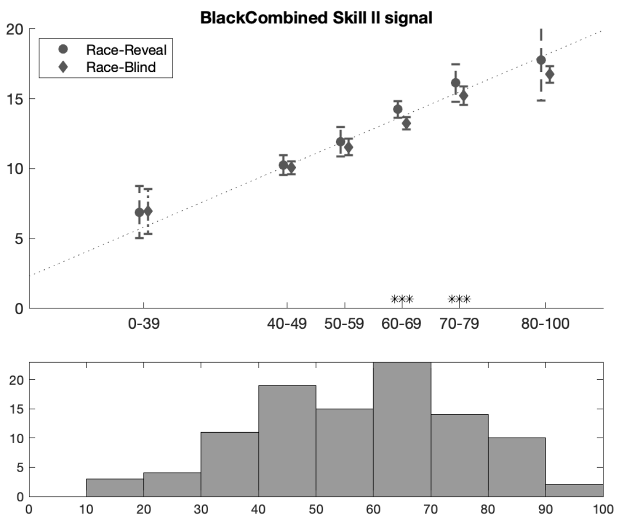
(C) Asian combined task hh signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

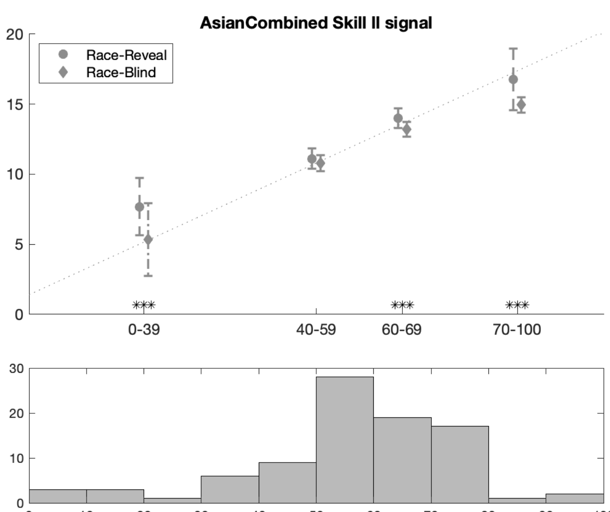
FIGURE 9. Combined task hh signal



(A) White combined task ll signal



(B) Black combined task ll signal



(C) Asian combined task ll signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

FIGURE 10. Combined task ll signal

It is a bit unclear whether the hl and lh signals should be considered high or low signals. We suggest the result of hl first. The results show that Asians are never overpaid, but underpaid in the 0-79% range, which contains 72.34% of observations. They are underpaid by 6.07% on average in this range. Black people are underpaid in the 30-49% range by 3.95%, and 11.83% of observations are in this range. They receive a fair wage otherwise. Whites are underpaid in the 0-59% range by 13.18% and 8.42% of observations are in this range. They are overpaid in the 70-100% range by 9.07%, where 46.32% of observations lie. Thus, we can conclude that Asians are underpaid the most when revealing their racial identity. Black people are also underpaid, but both the degree and portion are much smaller compared to Asians. Whites also have an underpaid region, though almost half of White workers are overpaid when their race is revealed.

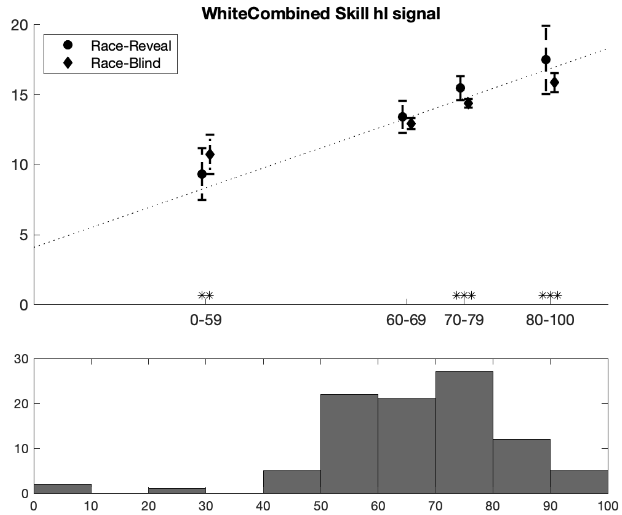
For a lh signal, Whites and Asians are never underpaid, while Black people are underpaid in the 0-49% range. In addition, Asians get more relative benefits compared to Whites. For Whites, 77.27% are in the overpaid range with 11.46% average overpayment, while the numbers are 91.94% and 16.01% for Asians. For Black people, 18.57% are underpaid by 15.31%, and the rest are overpaid by 5.85%. Thus, Black people are the only race who are underpaid when revealing their race.

Even if it is unclear whether to interpret hl and lh as high or low in general, we still find that discrimination exists against Asians and Black people in each case.

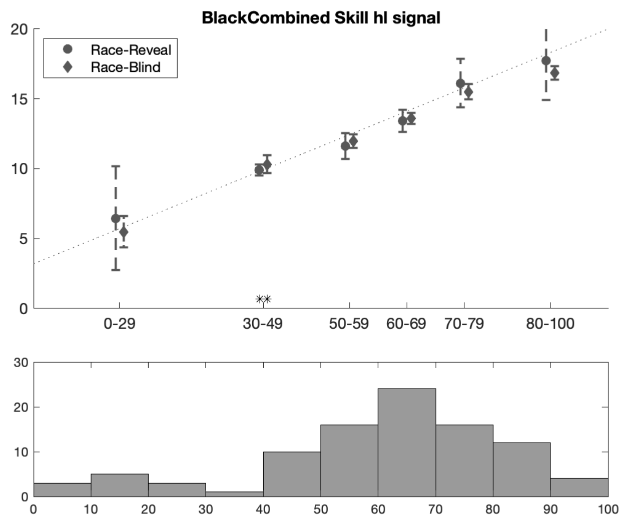
The results are demonstrated in Figure 11 and Figure 12.

Result 7. Discrimination in the combined task is similar to the math and social tasks: For certain ranges of beliefs, managers discriminate against Asians with hh and hl signals and Black people with a lh signal. Finally, all races get overpaid with a ll signal, but Black people get the least relative benefits.

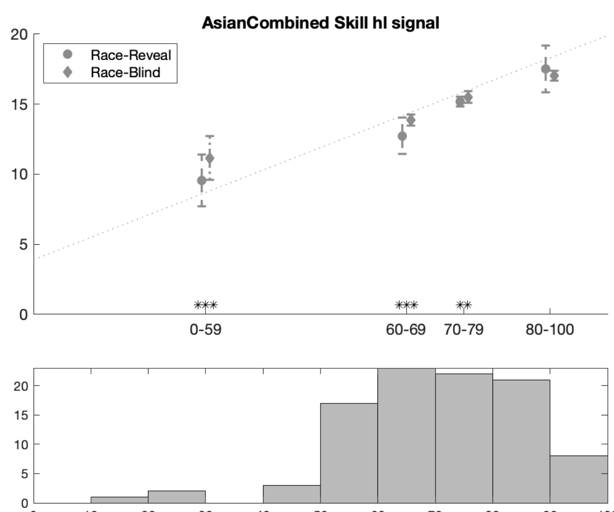
Also, this result suggests that if we focus on the first signal, the combined skill exhibits a similar pattern with math and social skills: an h signal gives a penalty to Asians while an l signal gives a penalty to Black people.



(A) White combined task hl signal



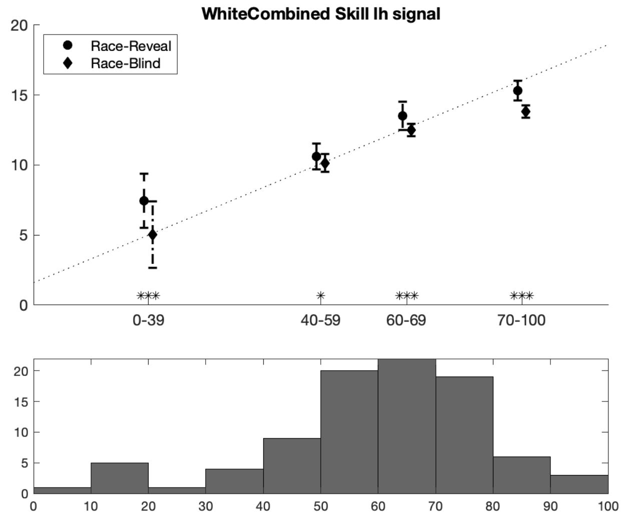
(B) Black combined task hl signal



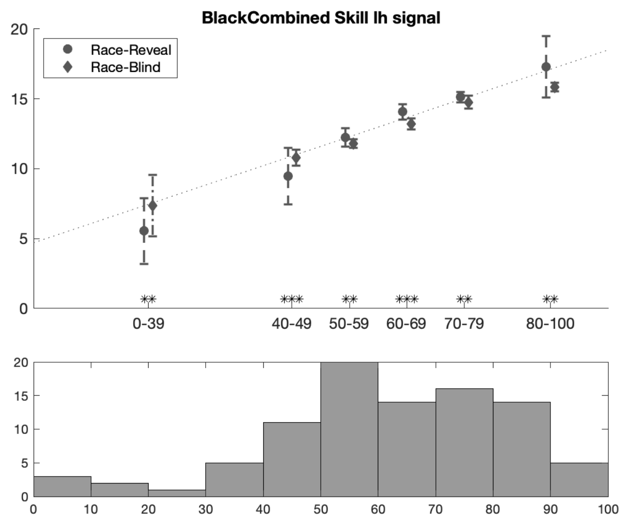
(C) Asian combined task hl signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

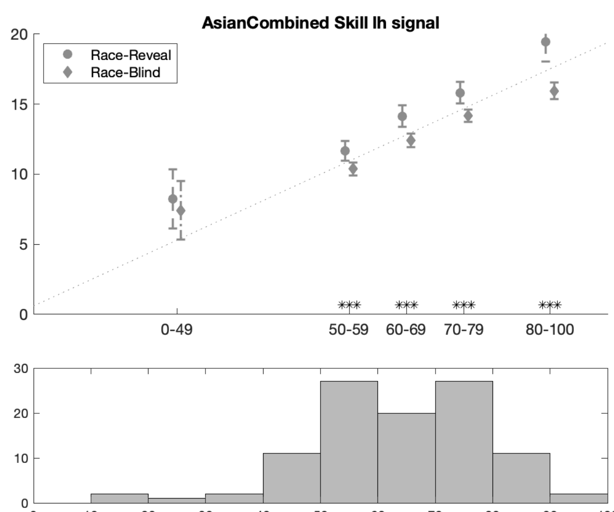
FIGURE 11. Combined task hl signal



(A) White combined task lh signal



(B) Black combined task lh signal



(C) Asian combined task lh signal

[1] * p<0.1, ** p<0.05, *** p<0.01.

FIGURE 12. combined task lh signal

6. ICT STUDY

After showing the existence of discrimination in Phase 2, we conduct an additional study to measure preferences for fixing discrimination. We adopt a method called Item Counting Technique, or ICT. ICT is used when a researcher wants to elicit responses to socially sensitive questions but is concerned about social desirability biases. Since discrimination is a socially sensitive issue, we choose the ICT experiment to measure the preference for fixing discrimination. The ICT method has been validated in similar settings by several previous papers (for example, Tourangeau and Yan, 2007, Blair and Imai, 2012, Blair et al., 2014).

6.1. *ICT design*

We follow directly the design of Coffman et al. (2017). There are four between-subjects treatments: Veiled-Asian, Direct-Asian, Veiled-Black, and Direct-Black. For convenience, a “veiled treatment" refers to both the Veiled-Asian and Veiled-Black treatments unless a race index is necessary. Likewise, a “direct treatment" refers to both the Direct-Asian and Direct-Black treatments.

In the veiled treatment, subjects are shown five statements that include one sensitive statement of interest. In our case, the sensitive statement is about their willingness to fix race discrimination. Subjects are asked how many of the statements correspond to them. In the direct treatment, subjects are shown the same four statements, excluding the sensitive statement, and then asked how many of the statements correspond to them. And then, separately, the sensitive statement is given as a form of yes/no question. Subjects are asked whether the answer is yes or no for them. We compare the responses from veiled treatment with the responses from direct treatment. For the direct treatment, a response is the summation of the response to the statements (from 0 to 4) plus the yes/no question (1 if yes, 0 if no). Then, the difference between the responses from the direct treatment and the veiled treatment captures the degree to which people do

Common Statements
1. I remember where I was during the January 6 riot at the United States Capitol.
2. I spent a lot of time playing video games as a kid.
3. I have voted for a political candidate who is known to be pro-life (opposed to abortion).
4. If I lived in a state that had not yet legalized recreational marijuana, and now there is a significant statewide referendum to legalize it, I would vote yes to legalize recreational marijuana.

Veiled Treatment
5. If I were in a team doing a peer review of a Black (Asian) co-worker, and others were unfairly giving them a low score on their quantitative (people) skills, I would try to compensate by giving a higher score.

Direct Treatment
Yes/No Question
If you were in a team doing a peer review of a Black (Asian) co-worker, on their quantitative (people) skills, and others were giving them an unfairly low score, would you try to compensate by giving a higher score?

TABLE 5. ICT Questions

not want to reveal their true response to the sensitive question. The specific statements used are provided in Table 5.

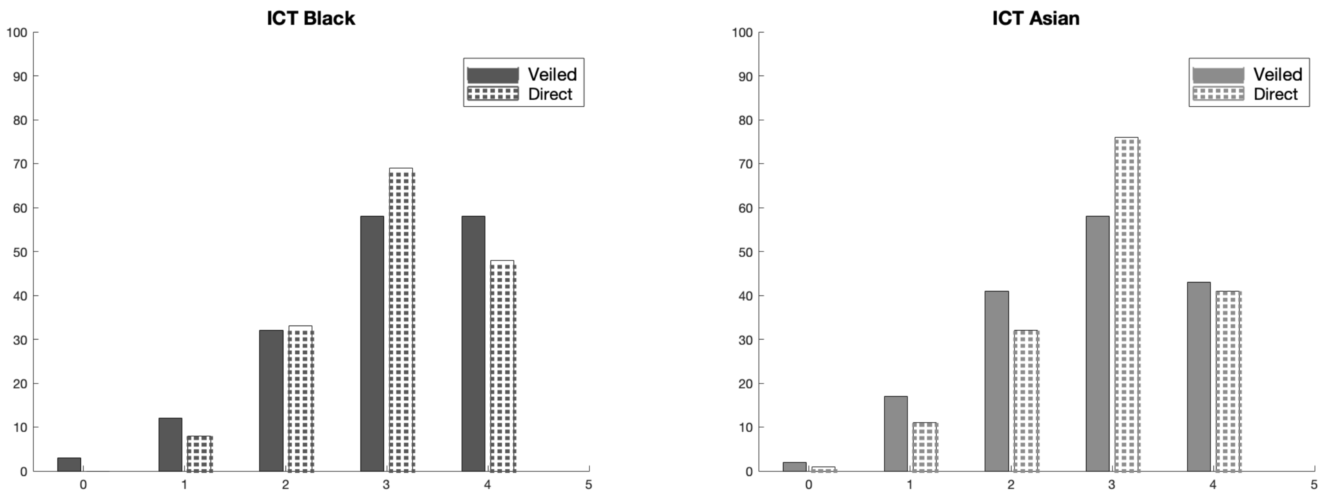
We make minor changes in Coffman et al. (2017)'s four insensitive statements and replace sensitive items with those of our interest regarding race discrimination.[15]

We provided an attention check to ensure subjects read the instructions. We recruited 643 subjects (322 for the Asian treatments and 321 for the Black treatments). The experiment lasted 3 minutes on average, and subjects were paid only upon completion. The completion fee is $1.

6.2. *ICT Results*

In this subsection, we show the results from the ICT study, which allows us to measure preference for fixing discrimination. Figure 13a and Figure 13b show the ICT results for both Black people and Asians. Each figure illustrates a histogram of the reported

[15]One of their statements includes the day of the *Challenger Space*. We worry that the current Prolific subjects are too young to know about this event, so we replaced it with the January 6 riot. Also, we make several statements longer to avoid our sensitivity statement being noticeable because of its length.



(A) ICT Result for Black (B) ICT Result for Asian

FIGURE 13. ICT Results

numbers, from 0 to 5. For the direct treatment, the answer to the yes/no question is included, with an additional count for the "yes" response. A "no" response indicates that a subject is not willing to fix discrimination. Therefore, a higher number suggests a more positive attitude towards fixing discrimination.

For the direct yes/no questions, 30.43% of subjects in the Asian treatment and 29.45% of subjects in the Black treatment respond with *no*. That means more than a quarter of people do not want to be actively engaged in fixing discrimination, even if it is obviously visible.

Another finding to notice is that there are no significant differences between the direct treatment and veiled treatment responses, both for Asians and Black people. Specifically, for Asians, the average responses are 2.81 for the veiled treatment and 2.93 for the direct treatment. Similarly, for Black people, the average responses are 3.03 for the veiled treatment and 3.04 for the direct treatment. This no-difference result suggests that people are comfortable revealing their preference for fixing or not fixing discrimination.

Finally, we test whether there is a difference between the discrimination-fixing preference for Asians and Black people. For this purpose, we compare the veiled treatments from the two racial groups since the veiled treatments are assumed to represent more honest responses. A marginal difference exists between Asians and Black people with a p-value of 0.069. Even if the p-value is greater than 0.05, considering that the ICT is a low-powered test, we can still take the difference (7.83% difference) as somewhat significant. This result suggests that people are somewhat more accepting of not fixing discrimination against Asians.

Result 8. Around 30% of subjects claimed they would not try to compensate for observed discrimination in the workplace. Furthermore, they reveal this truthfully even when the question is not veiled.

Result 9. Respondents are marginally less willing to fix discrimination against Asians compared to discrimination against Black people.

7. Discussion

We use a resume evaluation experiment to show that discrimination against both Asians and Black people exists and that the patterns differ between races. With our design, we contribute to discrimination studies in several ways. First, we consider multiple minority groups. There are cases where more than two minority groups exist, and they are treated differently. Most of the existing studies consider only one minority, whether choosing only one minority among all, or pooling all minority groups into one group. This approach could lead to an underestimation of discrimination since it may miss heterogeneity in discrimination across different minorities. Second, our use of a Race-Blind control allows us to separate out the causal effect of race on wages. The comparison between the Race-Revealed and Race-Blind treatments enables us to identify the effect of race on taste-based discrimination. Our main results suggest that there are differences in discrimination patterns: Black people with low signals and Asians with high signals receive lower wages. In some cases, Whites are not the most favored group. However, there are no cases in which Whites are harmed by race information more than either minority group. This result has a policy implication that there can be no one-size-fits-all policy that is appropriate for every minority group. Each minority would need a different policy to be protected from discrimination.

For example, consider the recent decision of the Supreme Court regarding affirmative action. The justices banned affirmative action in college admissions because affirmative action allegedly harms Asians by giving quotas to Black and Hispanic students. However, as we observed from our experiment, the domains in which Asians and Black people are discriminated against are different. Black people are discriminated against with low signals, which may justify why they benefit from affirmative action: At the margin of college admissions, they might be discriminated against due to their race. This can lead to the under-representation of Black people, and affirmative action may help offset this discrimination. On the contrary, Asians are discriminated against when their resume provides a high signal, so the effect of affirmative action for Asians may be small. They are thought to be over-represented, but perhaps not as much as they should be. Further, these discrimination patterns fit pervasive stereotypes. If we interpret signals as a measure of the level of diligence, then we can conclude that people discriminate against minorities whose resumes exemplify existing stereotypes of their racial group identity.

Our experiment is only designed to identify evidence of discrimination but not underlying mechanisms that cause such preferences. This may be an interesting avenue for future research. For example, we could investigate why discrimination does not occur at extremely high beliefs. We may also be able to study where the heterogeneity of beliefs for the same quality of resumes comes from, and whether this is due to selection or preference. Another direction would be to examine the relationship between managers' and workers' racial identities, such as in-group or out-group biases. From the previous literature, Fershtman and Gneezy (2001) documented that ethnic discrimination only comes from males, and both majority and minority male groups discriminated against minority male groups. Or, in a totally abstract setting, Chen and Li (2009) shows people exhibit in-group bias. Following those findings, it would be interesting to figure out how such biases operate in the context of race discrimination.

References

Aaronson, D., Phelan, B.J., 2022. The evolution of technological substitution in low-wage labor markets. Review of Economics and Statistics , 1–45.

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

Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. The Quarterly Journal of Economics 115, 715–753.

Alesina, A., La Ferrara, E., 2014. A test of racial bias in capital sentencing. American Economic Review 104, 3397–3433.

Altonji, J.G., Pierret, C.R., 2001. Employer learning and statistical discrimination. The Quarterly Journal of Economics 116, 313–350.

Antonovics, K., Knight, B.G., 2009. A new look at racial profiling: Evidence from the boston police department. The Review of Economics and Statistics 91, 163–177.

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

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

Arai, M., Thoursie, P.S., 2009. Renouncing personal names: An empirical examination of surname change and earnings. Journal of Labor Economics 27, 127–147.

Arnold, D., Dobbie, W., Hull, P., 2022. Measuring racial discrimination in bail decisions. American Economic Review 112, 2992–3038.

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

Arrow, K.J., 1974. THE THEORY OF DISCRIMINATION. Princeton University Press, Princeton. pp. 1–33. doi:doi:10.1515/9781400867066-003.

Avenancio-León, C.F., Howard, T., 2022. The assessment gap: Racial inequalities in property taxation. The Quarterly Journal of Economics 137, 1383–1434.

Babcock, L., Recalde, M.P., Vesterlund, L., Weingart, L., 2017. Gender differences in accepting and receiving requests for tasks with low promotability. American Economic Review 107, 714–747.

Badgett, M.L., Carpenter, C.S., Sansone, D., 2021. Lgbtq economics. Journal of Economic Perspectives 35, 141–170.

Bartoš, V., Bauer, M., Chytilová, J., Matějka, F., 2016. Attention discrimination: Theory and field experiments with monitoring information acquisition. American Economic Review 106, 1437–75.

Becker, G.S., 1971. The Economics of Discrimination. University of Chicago Press Economic Books.

Bertrand, M., Mullainathan, S., 2004. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. American Economic Review 94, 991–1013.

Blair, G., Imai, K., 2012. Statistical analysis of list experiments. Political Analysis 20, 47–77.

Blair, G., Imai, K., Lyall, J., 2014. Comparing and combining list and endorsement experiments: Evidence from afghanistan. American Journal of Political Science 58, 1043–1063.

Blau, F.D., Kahn, L.M., 2017. The gender wage gap: Extent, trends, and explanations. Journal of economic literature 55, 789–865.

Bohren, J.A., Haggag, K., Imas, A., Pope, D.G., 2019a. Inaccurate Statistical Discrimination: An Identification Problem. Working Paper 25935. National Bureau of Economic Research.

Bohren, J.A., Imas, A., Rosenberg, M., 2019b. The dynamics of discrimination: Theory and evidence. American Economic Review 109, 3395–3436.

Booth, A.L., Leigh, A., Varganova, E., 2012. Does ethnic discrimination vary across minority groups? evidence from a field experiment. Oxford Bulletin of Economics and Statistics 74, 547–573.

Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2016. Stereotypes*. The Quarterly Journal of Economics 131, 1753–1794.

Burgess, S., Greaves, E., 2013. Test scores, subjective assessment, and stereotyping of ethnic minorities. Journal of Labor Economics 31, 535–576.

Chan, A., 2023. Discrimination against doctors: A field experiment.

Charles, K.K., Guryan, J., 2011. Studying discrimination: Fundamental challenges and recent progress. Annual Review of Economics 3, 479–511.

Chen, Y., Li, S.X., 2009. Group identity and social preferences. American Economic Review 99, 431–57.

Christensen, P., Timmins, C., 2022. Sorting or steering: The effects of housing discrimination on neighborhood choice. Journal of Political Economy 130, 2110–2163.

Coffman, K.B., Coffman, L.C., Ericson, K.M.M., 2017. The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. Management Science 63, 3168–3186.

Esponda, I., Oprea, R., Yuksel, S., 2023. Seeing what is representative.

Exley, C.L., Kessler, J.B., 2022. The gender gap in self-promotion. The Quarterly Journal of Economics 137, 1345–1381.

Exley, C.L., Nielsen, K., 2023. The gender gap in confidence: Expected but not accounted for.

Eyting, M., 2023. Why do we discriminate? the role of motivated reasoning.

Feigenberg, B., Miller, C., 2022. Would eliminating racial disparities in motor vehicle searches have efficiency costs? The Quarterly Journal of Economics 137, 49–113.

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

Frankel, A., 2021. Selecting applicants. Econometrica 89, 615–645.

Fryer, R.G., 2007. Belief flipping in a dynamic model of statistical discrimination. Journal of Public Economics 91, 1151–1166.

Fryer, R.G., 2019. An empirical analysis of racial differences in police use of force. Journal of Political Economy 127, 1210–1261.

Goldsmith, A.H., Hamilton, D., Darity Jr, W., 2006. Shades of discrimination: Skin tone and wages. American Economic Review 96, 242–245.

Goncalves, F., Mello, S., 2021. A few bad apples? racial bias in policing. American Economic Review 111.

Haaland, I., Roth, C., 2023. Beliefs about racial discrimination and support for pro-black policies. Review of Economics and Statistics 105, 40–53.

Harpalani, V., 2022. Asian americans, racial stereotypes, and elite university admissions. Boston University Law Review 102, 233–326.

Heap, S.P.H., Zizzo, D.J., 2009. The value of groups. The American Economic Review 99, 295–323.

Holzer, H.J., Ihlanfeldt, K.R., 1998. Customer discrimination and employment outcomes for minority workers. The Quarterly Journal of Economics 113, 835–867.

Kessler, J.B., Low, C., Sullivan, C.D., 2019. Incentivized resume rating: Eliciting employer preferences without deception. American Economic Review 109, 3713–44.

Kline, P., Rose, E.K., Walters, C.R., 2022. Systemic Discrimination Among Large U.S. Employers*. The Quarterly Journal of Economics .

Kline, P., Walters, C., 2021. Reasonable doubt: Experimental detection of job-level employment discrimination. Econometrica 89, 765–792.

Kreisman, D., Rangel, M.A., 2015. On the blurring of the color line: Wages and employment for black males of different skin tones. Review of Economics and Statistics 97, 1–13.

Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? Journal of the European Economic Association 11, 495–524.

Lang, K., Spitzer, A.K.L., 2020. Race discrimination: An economic perspective. Journal of Economic Perspectives 34, 68–89.

List, J.A., 2004. The nature and extent of discrimination in the marketplace: Evidence from the field. The Quarterly Journal of Economics 119, 49–89.

List, J.A., 2006. Friend or foe? a natural experiment of the prisoner's dilemma. The Review of Economics and Statistics 88, 463–471.

Mayovich, M.K., 1972. Stereotypes and racial images —white, black and yellow—. International Journal of Social Psychiatry 18, 239–253.

McDonald, J.F., Moffitt, R.A., 1980. The uses of tobit analysis. The Review of Economics and Statistics 62, 318–321.

Mengel, F., Campos Mercade, P., 2022. Non bayesian statistical discrimination.

Miller, J.D., 1984. A new survey technique for studying deviant behavior. The George Washington University.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? do men compete too much? The quarterly journal of economics 122, 1067–1101.

Ondrich, J., Ross, S., Yinger, J., 2003. Now you see it, now you don't: why do real estate agents withhold available houses from black customers? Review of Economics and Statistics 85, 854–873.

Phelps, E.S., 1972. The statistical theory of racism and sexism. The American Economic Review 62, 659–661.

Reuben, E., Sapienza, P., Zingales, L., 2014. How stereotypes impair women's careers in science. Proceedings of the National Academy of Sciences 111, 4403–4408.

Shi, Y., Zhu, M., 2023. "model minorities" in the classroom? positive evaluation bias towards asian students and its consequences. Journal of Public Economics 220, 104838.

Szymanski, S., 2000. A market test for discrimination in the english professional soccer leagues. Journal of political Economy 108, 590–603.

Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. Psychological bulletin 133, 859.

Wood, P.B., Chesser, M., 1994. Black stereotyping in a university population. Sociological Focus 27, 17–34.

Wozniak, A., 2015. Discrimination and the effects of drug testing on black employment. Review of Economics and Statistics 97, 548–566.

Åslund, O., Hensvik, L., Skans, O.N., 2014. Seeking similarity: How immigrants and natives manage in the labor market. Journal of Labor Economics 32, 405–441.

Appendix A. Experimental Design Details

A.1. *Phase 1: The Worker Study*

General Instruction

- At first, you will be asked to provide demographic information:
  - 1. Hometown, 2. number of AP and/or GI classes you took in high school, 3. number of involvement/extracurricular activities during high school, 4. MS Word skill, and 5. Adobe Illustrator skill.
  - If you do not want to provide this information, please click "Leave". In this case, you will be paid nothing.

    Leave
- And then you will have three tasks.
  - You will solve math problems or make a decision in an economic situation in each task.
  - One of the tasks will be chosen for bonus payment.
  - Each point earned in each round will be converted to 10 cents.

Next

Demographic Information

- What is your hometown?
- What is your MS Word proficiency level? --Please choose an option--
- What is your Adobe Illustrator proficiency level? --Please choose an option--
- How many AP and/or GI classes did you take in high school (approximately)?
- How many involvement/extra curricular activities did you have in high school (approximately)? (If you are unsure about the exact number, please put the approximate.)

Next

Task 1 of 3: Instructions

- You will have 10 multiple-choice math questions.
  - For each question, you will have up to 15 seconds to solve.
  - You can pass any question but cannot come back to that question.
  - You will get 1 point per correct question.
  - Questions that are passed or unfinished will earn 0 point.

- We are interested in determining how many of these math questions you can get right without any help.
- So please do not use a calculator or look up the answers online, but rather just do your best.

Next

Task 2 of 3 Instructions (1/2)

- In this task, you will evaluate whether a decision maker's (called Ann) behavior is socially appropriate.
  - "Socially appropriate" means most people think the behavior is a "correct" or "ethical" thing to do.
- The following is the specific scenario you will consider.
  - There are two people, Ann and Bob. Each has $5 and they do not know each other.
  - Ann can decide whether to take money from Bob or give money to Bob.
  - The giving/taking amount can be from $0 to $5.
- Your job is to decide how socially appropriate each giving/taking behavior would be.

Please go to the next page for more detailed explanations and the payment scheme.

Next

Task 2 of 3 Instructions (2/2)

- The following is a screenshot of an example.

Take $5 (Ann,Bob) =($10,$0)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate

- On the left is Ann's behavior and the resulting payoffs.
- Your job is to choose the social appropriateness of each behavior, without knowing what others choose.
  - Recall that social appropriateness is determined by what most people think.
  - For each behavior, you will get points only if your choice is the same as the one that is most often chosen by others.

Payment

- If this task is chosen for your payment, one of Ann's behaviors is randomly chosen.
- If your response for that behavior is the same as that most frequently given by other people, then you will get 10 pts. Otherwise, you will get nothing.
  - *Note that it will take time to gather all participants' decisions. You will be paid the participation fee within 24 hours after finishing all tasks, and additional payment will be paid within 72 hours.*

Comprehension Check

- Which one of the following is **False**?

○ My job is to choose the social appropriateness of each behavior.
○ Payment is determined by one randomly chosen behavior.
○ You get 10 pts if your choice matches more than half of others' choices.

Next

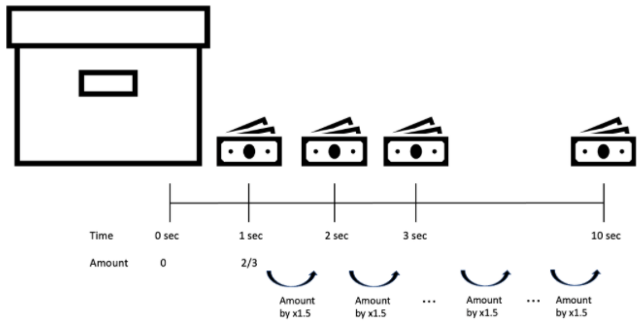
Task 2 of 3

- If you want to see the details again, click the following button.
  Scenario Details
- For each behavior of *Ann*, please choose the social appropriateness.

Take $5 (Ann:$10 Bob:$0)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Take $4 (Ann:$9 Bob:$1)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Take $3 (Ann:$8 Bob:$2)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Take $2 (Ann:$7 Bob:$3)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Take $1 (Ann:$6 Bob:$4)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Take/Give $0 (Ann:$5 Bob:$5)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Give $1 $0 (Ann:$4 Bob:$6)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Give $2 (Ann:$3 Bob:$7)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Give $3 (Ann:$2 Bob:$8)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Give $4 (Ann:$1 Bob:$9)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate
Give $5 (Ann:$0 Bob:$10)	very socially inappropriate	somewhat socially inappropriate	somewhat socially appropriate	very socially appropriate

Task 3 of 3 Instructions (1/2)

- Consider the following situation.

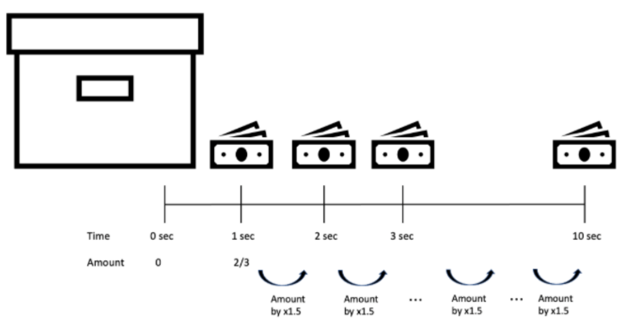


- There is a box that starts off empty, but every second, an increasing number of points is added to it.
  - The accumulation of points continues for a total of 10 seconds.
  - At one second, the box contains 2/3 of points.
  - After that, the number of points in the box is multiplied by 1.5 every second.
- There are two players who divide the money in the box.
  - Each player chooses when to stop the clock and take the money (points) currently in the box.
  - The one who stops first will get 80% of the total money in the box at that time, and the other one gets the remaining 20%.
  - If both stop at the same time, they will equally divide the money in the box.

Please go to the next page for more detailed explanations and the payment scheme.

Next

Task 3 of 3 Instructions (2/2)



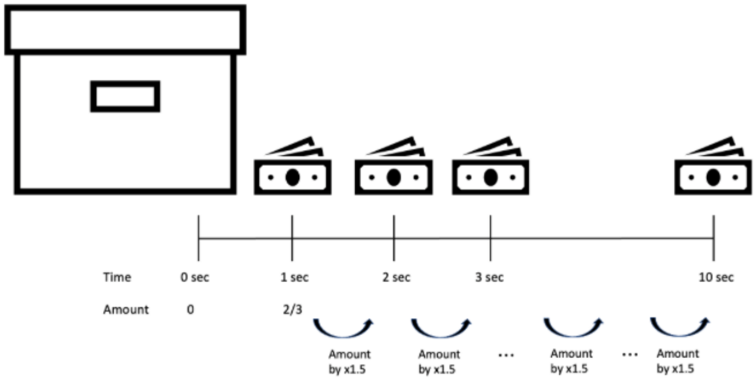
- You will actually play this game against every other person who participates in this study.
  - For example, if there are 100 other participants, you will play the game 100 times.
- But you choose when to stop only once, and that choice gets used for all 100 games.
- If this task gets chosen for your payment, your payoff will be the average payoff from all games.
  - *Note that it would take time to gather all participants' decisions. You will be paid the participation fee within 24 hours after finishing all tasks, and additional payment will be paid within 72 hours.*

Comprehension Check

- Which one of the following is **True**?
  - ○ I will play this game only once, with one other participant.
  - ○ I will choose my stopping time multiple times.
  - ○ My payment will be average across all games.

Next

Task 3 of 3



- The one who stops first will get 80% of the total money in the box at that time, and the other one gets the remaining 20%.
- If both stop at the same time, they will equally divide the money in the box.

When do you want to stop? Please choose between 0 and 10 seconds.

After 115 seconds you can move to the next page whenever you are ready.
If you do not move to the next page, your answer will not be recorded, and you will not get paid.

A.2. *Phase 2: The Evaluator Study*

Instructions (1/2): Tasks

- We ran a previous study to measure the abilities of 282 workers.
  - The workers performed three tasks that measure
    (1) Math skill, (2) Social skill, and (3) Combined (Math+Social) skill.
  - If you want to know the details of tasks, click the following buttons.
    Math skill **Social skill** **Combined (Math+Social) skill**
  - For each ability, a worker is one of two types: High or Low.
    - Types for each task are defined as,

Task	Low	High
Math	< 50% correct	≥ 50% correct
Social	< 50% of points	≥ 50% of points
Combined	< 1.6 pts	≥ 1.6 of points

  - The workers are US citizens, currently living in the US now.
- Your task is to evaluate 14 workers' abilities from that study.

- The evaluation task consists of two parts:
  - For the first part, you will report your beliefs about 10 workers' abilities for all three skills.
    - For each worker, you will report 3 beliefs: Probability of High type for each skill (Math, Social, Combined).
  - For the second part, you will report what wages you'd be willing to pay to 4 different workers for all three skills.
    - The workers may actually get paid, based on your decision.
    - You will see the details in the payment instructions.
- There will be attention check questions throughout the study.

Comprehension Check

- Which one of the following is **True**?
  - ○ Some of the workers do not live in the US or are not US citizens.
  - ○ The combined skill task requires both math and social skills.
  - ○ For each worker, only one skill will be evaluated.

Instructions (2/2) : Payment Scheme

- One of 14 workers will be randomly chosen for your payment.
- Each point earned will be converted to 10 cents.
- If a worker is chosen from the first task, your payoff is determined as follows:
  - One of your belief on three tasks (Math, Social, or Combined) will be randomly chosen.
  - You may win 20 pts based on your answer.
  - We use a formula to calculate your payment. Click **here** to see details.
  - With this formula, your payment is **maximized** when you report belief **truthfully.**
- If a worker is chosen from the second task, your payoff is determined as follows:
  - We randomly pick one of the three tasks (Math, Social, or Combined).
  - You get paid a "revenue" based on their type on that skill (High or Low).

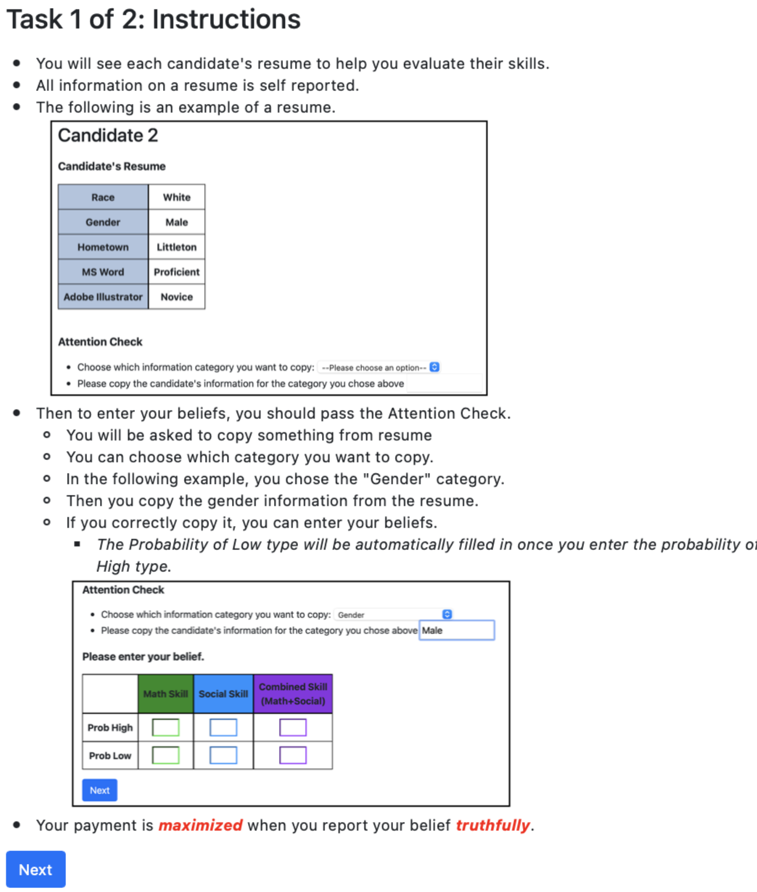
Type	Revenue
High	20 pts
Low	10 pts

  - For each skill, you submit the maximum wage you're willing to pay.
  - The computer randomly draws an actual wage (from 0 to 20).
  - If the actual wage is less than your maximum wage then you hire the worker for that skill
    - You receive the revenue based on the table above.
    - You pay the actual wage drawn by the computer.
    - You keep the difference between the revenue and the actual wage.
    - The worker will receive the actual wage.
    - If you earn negative points, your earning will be automatically adjusted to 0.
  - If the actual wage is above your maximum wage then you will not hire the worker for that skill
    - In this case both you and the worker earn nothing.

Comprehension Check

- Which one of the following is **True**?
  - Your choice will affect a worker's payoff whoever is chosen.
  - Multiple workers will be chosen for your payment.
  - If a worker is chosen from the second part, there is a chance that a worker is paid.

- Consider the wage part. Suppose that your maximum wage is 12. Which of the following is **False**?
  - If the actual wage is 8, you will pay 8 to the worker.
  - If the actual wage is 11, and the worker's type is Low, both you and worker will earn 0.
  - If the actual wage is 13, you will get nothing.



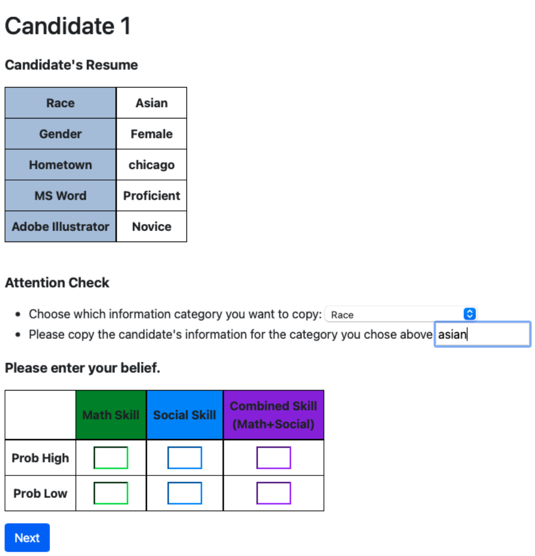


FIGURE 14. Prior: Race-Revealed

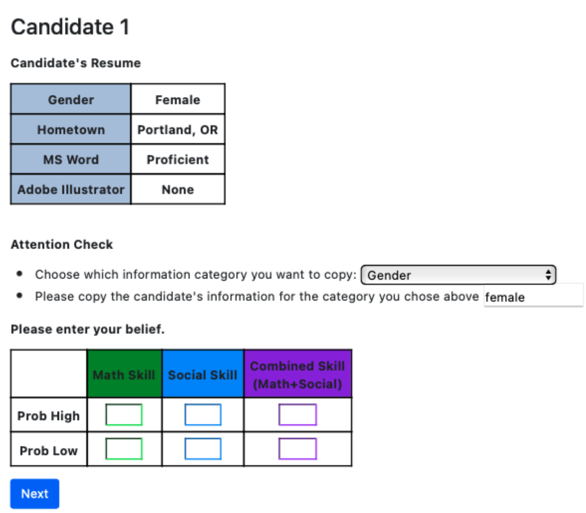


FIGURE 15. Prior: Race-Blind

Task 1 of 2

- For the remaining four candidates, you will see two additional items on their resume: AP/IB classes and Extracurricular activities.
  - Again, all information on a resume is self reported.
  - "AP/GI classes" indicates number of AP/IB classes they took in high school.
  - "Extracurricular activities" indicates number of extracurricular activities they had in high school.
  - Both will be given in range.
  - Like the previous 6 candidates, your payment is ***maximized*** when you report your belief ***truthfully***.

Next

Candidate 9

Candidate's Resume

Race	White
Gender	Female
Hometown	Columbia
MS Word	Proficient
Adobe Illustrator	Novice

AP/IB classes	over 3
Extracurricular activities	over 2

Attention Check

- Choose which information criteria you want to type: Hometown
- Please type the candidate's information for the category you chose above columbia

Please enter your belief.

	Math Skill	Social Skill	Combined Skill (Math+Social)
Prob High			
Prob Low			

Next

Task 2 of 2: Instructions

- You will see four more resumes.
- You will submit your maximum wages for all three skills.
- Recall that if a worker is chosen from this task for your payment, there is a chance that a worker is paid.
- Everything else will be the same as Task 1.

Next

Candidate 11

Candidate's Resume

Race	White
Gender	Female
Hometown	Syracuse, NY
MS Word	Proficient
Adobe Illustrator	Novice

AP/IB classes	over 3
Extracurricular activities	over 2

Attention Check

- Choose which information criteria you want to type: Adobe Illustrator
- Please type the candidate's information for the criteria you chose above: novice

Please enter the maximum wage you're willing to pay.

Recall that the revenue from High is 20, and the revenue from Low is 10.

What is the maximum wage you're willing to pay to this candidate for the **math** task? 10

What is the maximum wage you're willing to pay to this candidate for the **social** task? 10

What is the maximum wage you're willing to pay to this candidate for the **combined** task?

Next

A.3. *Phase 3: ICT Study*

A.3.1. *Veiled Treatment*

Instructions

- You will be given five statements.
- You will answer how many of the statements apply to you.
  - For example, if statement 1 and statement 3 apply to you and the other two do not, your answer will be 2.
  - Notice that the researchers will not be able to determine which statements apply to you. They will only see the total number that apply to you

- Please remember (or write down) this code: BEAR
- We will ask you to recall it at the end of the experiment.
- *Notice that this code is different from the Prolific completion code.*

Next

Task

Five Statements

- I remember where I was during the January 6 riot on the United States Capitol.
- I spent a lot of time playing video games as a kid.
- I have voted for a political candidate who is known to be pro-life (opposed to abortion).
- If I lived in a state that had not yet legalized recreational marijuana, and now there is a significant statewide referendum to legalize it, I would vote yes to legalize recreational marijuana.
- If I were in a team doing a peer review of a Black co-worker, and others were unfairly giving them a low score on their quantitative skills, I would try to compensate by giving a higher score.

Please fill in the bubble that corresponds to the total number of statements above that apply to you.

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Please type the four-letter code you were given in the opening instructions:

Next

A.3.2. *Direct Treatment*

Instructions

- You will be given four statements.
- You will answer how many of the statements apply to you.
    - For example, if statement 1 and statement 3 apply to you and the other two do not, your answer will be 2.
    - Notice that the researchers will not be able to determine which statements apply to you. They will only see the total number that apply to you
- In addition, you will be given one question that can be answered yes, or no.

- Please remember (or write down) this code: BEAR
- We will ask you to recall it at the end of the experiment.
- *Notice that this code is different from the Prolific completion code.*

Next

Task

Four Statements

- I remember where I was during the January 6 riot on the United States Capitol.
- I spent a lot of time playing video games as a kid.
- I have voted for a political candidate who is known to be pro-life (opposed to abortion).
- If I lived in a state that had not yet legalized recreational marijuana, and now there is a significant statewide referendum to legalize it, I would vote yes to legalize recreational marijuana.

Please fill in the bubble that corresponds to the total number of statements above that apply to you.
○ 0 ○ 1 ○ 2 ○ 3 ○ 4

Yes/No Question

- If you were in a team doing a peer review or a Black co-worker on their quantitative skills, and others were giving them an unfairly low score, would you try to compensate by giving a higher score? ○ Yes ○ No

Please type the four-letter code you were given in the opening instructions:

Next

APPENDIX B. SIMULATION DETAILS

B.1. *Procedure*

The simulations are done in the following procedure.

(0) Fix a racial group, a signal, and a task. Let n_1 be a number of observations for posterior beliefs and n_2 be a number of observations for wage offers. Also, let $n = \min\{n_1, n_2\}$

(1) Set grids for β_0 and β_1, say B_0 and B_1.

(2) Draw one $\beta_0 \in B_0$ and $\beta_1 \in B_1$.

(3) From the posterior vector, randomly draw posteriors with replacement to make $(n_1 + n_2) \times 1$ vector.

(4) From the drawn posteriors, make a wage vector such that

$$w_i = \min\{\beta_0 + \beta_1 \times post_i + \epsilon_i, 20\} \quad \text{where } \epsilon_i \sim N(0,1)$$

(5) Randomly draw n wages and n posteriors independently, without replacement. Let a new wage vector and new posterior vector as w^N and p^N.

(6) Order w^N and p^N from low to high.

(7) With the pair (w^N, p^N), estimates $\hat{\beta}$ that minimizes mean squared error.

(8) Repeat (3) to (7) 100 times and mean of 100 estimates the final estimates.

(9) Repeat this procedure for all $\beta_0 \in B_0$ and $\beta_1 \in B_1$