Society for Economics Research in India
Working Paper Series

DO WORKERS DISCRIMINATE AGAINST THEIR OUT-GROUP EMPLOYERS? EVIDENCE FROM AN ONLINE PLATFORM ECONOMY

Sher Afghan Asad
(Lahore University of Management Sciences, Lahore)
Ritwik Banerjee
(Indian Institute of Management, Bangalore)
Joydeep Bhattacharya
(Iowa State University, Ames)

Working Paper No. 5
http://seri-india.org/research

September 2020

Do workers discriminate against their out-group employers? Evidence from an online platform economy

Sher Afghan Asad
Lahore University of Management Sciences
sherafghan@lums.edu.pk

Ritwik Banerjee
IIM, Bangalore
ritwikbanerjee@iimb.ac.in

Joydeep Bhattacharya
Iowa State University
joydeep@iastate.edu

Abstract

We study possible worker-to-employer discrimination manifested via social preferences. We run a well-powered, model-based experiment, wherein we recruit 6,000 white American workers from Amazon's M- Turk platform for a real-effort task. We randomly (and unobtrusively) reveal the racial identity of their non-fictitious employer, who may either be white or black. We find evidence of race-based altruism towards black employers. However, the workers display significant racial discrimination in reciprocity - a small gift induces workers to put higher effort for white employers relative to black. Our results suggest that taste-based discrimination favoring ingroup can have significant adverse effects on outgroup employers.

Keywords: Discrimination; Worker-to-Employer; Social Preferences; Taste-based discrimination; Online Economy; Mechanical Turk; Structural Behavioral Economics.
JEL Codes: J71, D91, C93

We thank, without implicating, Martin Abel, Mackenzie Alston, Otavio Bartalotti, Michael Best, Kerwin Charles, Stefano DellaVigna, Catherine Eckel, Betsy Hoffman, Alex Imas, Peter Orazem, Devin Pope, Joshua Rosenbloom, Arka Roy Chaudhuri, and participants at the 15th ISI-Delhi Annual Conference, the Advances with Field Experiments (AFE, 2019) conference in Chicago, and the National Economic Association 2020 conference in San Diego for their invaluable input. We thank Bernard Fay and Alejandro Ruiz Ortega for excellent research assistance. We also acknowledge partial funding from the Russell Sage Foundation and from the Economics Department at Iowa State University. This research was approved by the IRB at Iowa State University, and was registered on the AEA RCT registry, ref. AEARCTR-0003885.

1 Introduction

By construction, *Homo economicus* is self-interested and only takes actions that maximize his/her payoffs. By way of contrast, *Homo behavioralis*, in addition to being self-interested is also endowed with social preferences, a concern for how his/her actions affect the payoffs of others. These "others" could belong to his in-group, a group he identifies with, and whose membership gives him a sense of belonging. Everyone else, by definition, is in his out-group. *Homo behavioralis* may harbor negative social preferences urging him to discriminate against others; or the preferences could be positive and take the form of prosocial behavior – actions taken with an intent to benefit others with no expectation of personal benefit.

This paper is aimed at detecting evidence of positive or negative social preferences within the context of labor markets. The experimental setting is a U.S. based online platform labor market, and group identity is assumed to be racial in origin. Within this environment, we ask, is there evidence that whites systematically treat blacks differently from how they treat fellow whites? We depart from a half-century of research in labor economics that views this issue largely as unidirectional, emanating from employers, and directed toward their employees.[1] Instead, we ask, is there evidence that white *workers* in the online economy treat their black *employers* better or worse than how they treat their otherwise-identical, white employers?

A series of questions come up right away. Why is it interesting to study discrimination or pro-social behavior of workers toward employers? Is there any evidence of this? And why the online economy? We take these up one by one. That workers may treat their out-race employers differently may, at first glance, appear implausible; after all, it is mostly bosses who get to frame labor contracts, and it seems within the bounds of such contracts, they will not leave much room to be mistreated. Our view is that this

[1] See Riach and Rich (2002), Charles and Guryan (2011), Rich (2014), Bertrand and Duflo (2017), and Neumark (2018) for a review of this literature.

first-pass line of thinking is limited. While admittedly it is easier for bosses to maltreat out-group workers, the latter are also keenly aware that the effort they put in, the diligence or care they show on the job, crucially affects the bottom line of their bosses. Moreover, as is well known, labor contracts are often "incomplete": they leave workers a considerable degree of discretion over work effort. It is therefore conceivable that a worker with substantial leeway over effort makes very different effort choices reflecting his underlying differential social preferences. For instance, a black worker may choose to work harder for a black boss because of his desire to a) see his boss succeed even if it does not benefit him personally (**altruism** à la Simon (1993)), and b) return any respect or kindness he receives from his boss (**reciprocity** à la Akerlof (1982)).[2] In this study, we limit attention to these two forms of social preference.

Second, there is important evidence that workers care about the social identity of their bosses and differentially perform for in versus out-group employers. Sundstrom (1994), focusing on U.S. urban labor markets 1910-1950, notes "one of the most widely noted rules of the southern labor market was that blacks were not to supervise whites...[because it] would plainly invert the appropriate hierarchy" which meant "blacks were generally absent from supervisory positions". White employees simply did not wish to receive orders from (or work under) black supervisors. More recently, Glover, Pallais, and Pariente (2017) study whether discriminatory beliefs held by bosses directly affect minority workers' job performance in a real-world workplace. They investigate the performance of cashiers in a French grocery store chain, and find when "minority cashiers, but not majority cashiers, are scheduled to work with managers who are biased (as determined by an Implicit Association Test), they are absent more

[2] Any employment generally involves a contract describing the exchange of work tasks, remuneration, and other obligations. Besides, the employment relationship also implicitly encapsulates perceptions and beliefs regarding the terms and conditions of a reciprocal exchange agreement between the parties. Such reciprocal links form the bedrock of labor market exchange in any economy. For example, in exchange for loyalty to the firm, hard work, and beyond-the-call-of-duty involvement from workers, bosses may offer them implicit assurances of job security, career development, and flexibility to achieve work-life balance. The latter are 'gifts' – they are wildly popular, ubiquitous, and often, kept outside of the contractual arrangement.

often, spend less time at work, scan items more slowly, and take more time between customers." The upshot is, workers *do* adjust their effort based on the social identity of their bosses, and may perform better when paired with own-group managers than out-group ones.[3]

And why study this question in the confines of the online economy? An online labor market platform economy is one where independent workers are paid by the gig (i.e., for a task or a project) as opposed to the traditional economy where workers are paid a salary or hourly wage as part of a contract. One important distinction is that in the online economy, particularly of the digital-platform type, there is less scope for familiarity or closeness or repeated physical interactions between the employer and the employee; hence, physical associative distaste is unlikely to be activated (Rotemberg, 2006). This means, if we are to detect any race-based differences in social preferences (altruism or reciprocity) in our online economy setting, it will not be because of associative distaste or liking. Another critical difference is the vastly dissimilar "power dynamic" between worker and boss. In an online economy, workers retain a lot of power in the worker-employer relationship: they may shirk under a particular employer or easily switch employers without losing much "employment rent". This new power dynamic makes the online economy an ideal setting to study *worker-to-boss* discrimination, much more so than the conventional labor market setting.

To the end of answering our research question, we run a well-powered, AEA pre-registered, no-deception, model-based experiment using 6,000 white subjects from one of the largest online economy platforms: Amazon's Mechanical Turk (M-Turk). Specifically, our experimental design uses U.S. based white subjects from M-Turk (recruited as "workers") and black and white student subjects (recruited as "employers") from a major U.S. public university. The interaction between a worker and an employer

[3] Oh (2019) finds that 43% of Indian workers "refuse to spend ten minutes working on tasks associated with other castes, even when offered ten times their daily wage" suggesting the important role of social identity in determining work-related decisions.

is kept one-shot, as is typical in the online economy, so that confounding reputation effects (of the kind that naturally emerge in Glover et al. (2017)) do not enter. In the experiment, workers engage in a real-effort task for a pre-assigned, non-fictitious employer who, in some treatments, may be racially identifiable as black or white. The task entails a real utility cost of effort (unlike those using monetized costs in studies such as Charness, Rigotti, and Rustichini (2007); Fershtman and Gneezy (2001)) because it requires a worker to alternately press the 'a' and 'b' buttons on a keyboard for up to 10 minutes.[4] Workers do not get to select their employer but are free to decide how much effort to provide on the task (an 'incomplete contract' environment).[5] The worker's performance is measured by the number of times the buttons are alternately pressed, and the worker is informed (truthfully) of the payoff the employer will receive due to the worker's performance.

The design is tightly connected to a simple structural model à la DellaVigna, List, Malmendier, and Rao (2016) in which workers are assumed to have race-dependent, social preferences (altruism and reciprocity) towards their employer and maximize utility subject to the cost of effort. Inspired by Doleac and Stein (2013), these preferences are activated in some treatments by unobtrusively signaling the employer's race to the matched worker via skin color and voice: employer-subjects are videotaped while they

[4] The task is admittedly artificial in the sense that workers, in reality, do not routinely engage in such meaningless tasks. We offer three arguments for choosing such a task. First, we wanted the task to not require any special ability on the part of workers(as is often the case for M-Turk work); in that case, our results would be tainted by the unobservability of the underlying ability distribution attached to the task. Second, the task is exactly the one used in DellaVigna and Pope (2018) thereby facilitating comparisons across our paper and theirs, more so because they use workers from M-Turk (we restrict participation to U.S. workers, they don't). And third, however meaningless the task may be, it requires substantial focus and effort, both of which contribute to measurable actual earnings.

[5] Traditionally, economists understand discrimination in labor markets to arise in two main ways. Becker (1957) introduced the notion of taste-based discrimination postulating that discrimination exists because of a prejudice/animus towards the members of the disadvantaged group. On the other hand, Phelps (1972) and later Arrow (1973), theorized that discrimination might be statistical – an employer, lacking information about a job-seeker's productivity, forms beliefs about it based on the person's group identity and the aggregate productivity distribution of the group to which the person belongs. In our experiment, the employers do not get to make any strategic choices (such as wage offers, payments, minutes of work, work times, etc.). This eliminates most channels for statistical discrimination by workers.

read off a script explaining and demonstrating the task for the workers. The camera placement only captures the hand of the employer along with the movement of the fingers alternating 'a' and 'b' button presses. Other identifiers, such as the face, are not revealed. In the race-neutral treatments, gloves and other clothing hide the skin entirely. We make every effort to check that race, when revealed, is correctly perceived. In the experiment, we introduce a total of ten treatment variations. In the first four, we vary the piece rate (0, 3, 6, and 9 cents) to identify and estimate the cost-of-effort function. Here, the worker is not given any information about the existence of (a non-existent) employer; any earnings from his/her effort choices go entirely to the worker. In the remaining six treatments, the worker is not only made aware of being matched to a specific, actual employer, but the connection between his effort and the employer's payoff is made clear. Among these, the first set of three treatments aim to a) detect the baseline level of altruism towards the hidden race of the employer (altruism-neutral (AN)) and b) estimate race-specific altruism towards the revealed race of the employer (altruism-black (AB) and altruism-white (AW)). The final three treatments combine altruism with a small, $0.20 monetary gift from an employer with either a concealed race (altruism and reciprocity – neutral (ARN)) or a salient race (altruism and reciprocity - black (ARB) and altruism and reciprocity - white (ARW)). When "differenced out" from the respective altruism treatments, the altruism and reciprocity treatments help a) detect the baseline level of reciprocity towards an employer with concealed race and b) estimate the race-specific variations in reciprocity towards employers with salient races. In all, the ten treatments help us identify the cost-of-effort function and social preference parameters (altruism and reciprocity) of the structural model separately for neutral (hidden race), black, and white employers.

Our findings reported in terms of average effort by white workers are as follows. First, not surprisingly, incentives via piece rates have a strong, statistically-significant effect on effort. This observation lends credence to the idea that the MTurk population is generally representative of a typical labor force: MTurkers, like most workers, work harder when they receive higher wages. Second, we detect statistically significant evidence for altruism: workers put more effort when they know their work benefits an employer of unknown race (AN treatment) as compared to the piece rate 0-cent treatment where neither the worker nor the employer earns any payoff attributable to worker effort.

Strikingly, white workers are significantly more altruistic towards black employers than white employers. In addition to being statistically significant at the 2% level, the difference in effort provision is non-trivial. To see this, consider a baseline level of altruism, defined as the differential effort provided by white workers knowing their effort enhances the payoff of an unknown race employer *versus* their effort when the piece rate is 0-cent and no employer exists. The differential effort by white workers, knowing their effort enhances the payoff of a white vs. a black employer, is 75% of this baseline. The structural estimation exercise also reveals that black employers get 5% more effort than white employers at a 0-piece rate. Collectively, these represent persuasive evidence of positive altruism by white workers toward black employers. However, the story does not end there. When the results from the social-preference treatments are compared, we find no statistical difference between the average effort for white and black employers. This implies that any advantage black employers receive from positive altruism is essentially wiped away by the negative response to the gift exchange. In fact, the pure average effort response of white workers to the $0.20 monetary gift from a black employers virtually zero and significantly positive for a white employer. This indicates white workers reciprocate significantly less to a gift from a black relative to a white employer. This offers compelling evidence of a race-based, discriminatory response in reciprocity by white workers towards black employers.

We make several significant contributions to the literature. First, to our knowledge, this is the first paper to demonstrate that racially motivated, taste-based discrimination (in reciprocity) can emerge in workers and be directed toward employers. Our results work environments where reciprocal, worker-employer relationships dominate, such taste-based discrimination in favor of the ingroup can have a significantly adverse effect on outgroup employers. Our work relates to a small, emerging literature studying discrimination by subordinates (Abel, 2019; Ayalew, Manian, & Sheth, 2018; Chakraborty & Serra, 2019; Grossman, Eckel, Komai, & Zhan, 2019). This literature focuses on gender as group identity and mostly finds belief-based discrimination against female leaders. Another study on Amazon's Mechanical Turk Abel (2019) finds that workers do not discriminate in effort choices when they work for women leaders, even though the feedback from them is perceived as being less pleasant than from a male leader. Evidence from Benson, Board, and Meyer-ter Vehn (2019) suggests that workers' performance is influenced by the social identity of their boss. They chalk it to the fact that bosses can better screen applicants from their own race. Our design shuts down this "selection effect", which in turn allows us to investigate the possibility of race-based discrimination by subordinates in the U.S., which is mainly taste-based in origin.

Second, an important branch of the empirical economics literature on discrimination tries to document the effects of employer-to-worker discrimination on, say, wage gaps between black and white workers, those that remain after statistically controlling for observable characteristics of workers.[6] The unexplained differences are *assumed* to be the result of taste-based discrimination. As pointed out by Charles and Guryan (2011), these studies cannot get to the causal effect of being black or white on worker earnings because it is impossible to assign worker race randomly. Correspondence studies – see Bertrand and Duflo (2017) – get around this by randomly varying the *perception* of race. Inspired by this tradition, we estimate the causal effect of perceived employer race on worker effort. In particular, the treatment variations we study, along with

[6] See Riach and Rich (2002), Charles and Guryan (2011), Rich (2014), Bertrand and Duflo (2017), and Neumark (2018) for a review of this literature.

our tight connection to a structural model, allows us to separately identify unambiguous parameter estimates of race-dependent, discrimination in altruism and reciprocity. These findings represent important advances over the Oaxaca-Blinder-style decomposition exercises. The parameter estimates we find would be useful for conducting welfare analyses. It bears emphasis that our strategy of signaling race does not reveal anything more than race, a common critique of correspondence studies.

Third and relatedly, we identify discrimination not in the usual Beckerian sense, where animus affect extensive margin hiring decisions, but as a preference parameter which affects choice (effort, in our case) on the intensive margin. Our study shows that the intensive margin can be an important source of discrimination against minorities. This is important because the legal framework has traditionally focused attention solely on external-margin discrimination, and has mostly ignored discrimination on the intensive margin, the type that may emerge in incomplete contract settings.

Finally, our work is related to the larger topic of "what motivates effort?" especially in incomplete contract environments (Fehr & Gächter, 1998). It goes beyond DellaVigna and Pope (2018) to suggest, for example, that structurally-estimated, social preference parameters which govern the supply of effort ought to be indexed by race. Our work also connects to the literature on gift-giving. That literature has established the importance and ubiquity of reciprocal gift-exchange in the American workplace (Dodlova & Yudkevich, 2009; Ferrary, 2003). This literature mainly uses simple lab-based games such as the dictator game or the trust game to detect discrimination in altruism and reciprocity. Strikingly, it almost always finds that discrimination in altruism and reciprocity go hand in hand (Bertrand & Duflo, 2017; Fershtman & Gneezy, 2001). In contrast, we find that discrimination in altruism and reciprocity parameters can, in fact, operate in opposite directions.

The rest of the paper proceeds as follows. In Section 2, we present the model of behavior and produce the treatments to identify the parameters of interest. In Section

3, we present the experiment design. Section 4 summarizes the data. In Section 5, we present the results followed by structural estimation in Section 6; concluding remarks are in Section 7.

2 Model and Treatments

In this section, we present the model of behavior that is used to design the experiment. The model explains a worker's effort choice given the monetary and non-monetary incentives and costs of working for an employer. Our design is inspired by DellaVigna et al. (2016) modified to permit discrimination from the workers' side. In the setup, workers choose how much effort to provide on a real-effort task.

A risk-neutral worker, working for an employer j, $j \in \{Neutral, Black, White\}$, receives utility[7]

$$U_j \equiv (\ F + (s + \rho_j \mathbb{1}_{Gift} + \alpha_j v + p) e_j - c(e_j)). \tag{1}$$

Here, e_j is the number of points (on the button-pressing task) scored by the worker when working for an employer j, F is the fixed participation fee he receives , and s captures a sense of duty, norm, intrinsic motivation, and competitiveness of the worker towards the task and is independent of the employer. ρ_j is the reciprocity parameter per unit of effort which is activated whenever employer j awards a gift to the worker à la Gneezy and List (2006). $\mathbb{1}_{Gift}$ is an indicator function which assumes a value 1 when a gift is rewarded by the employer, 0 otherwise. α_j captures the altruistic preference of a worker towards employer j per unit of effort à la Becker (1974), where v is the (race independent and exogenous) value to the employer of a unit of effort by the worker. Note that our notion of altruism captures "pure altruism" as well as "warm glow" of the workers (DellaVigna et al. (2016)): we don't aim to disentangle the two. p is the piece rate per unit of effort. $c(e_j)$ is the cost of effort function, assumed, for now, to

[7] We assume risk neutrality because the stakes are too small for the curvature of the preferences to matter. It also leaves us with one less parameter to estimate.

be the same for all workers. We assume the regularity conditions $c'(.) > 0$, $c''(.) > 0$, and $lim_{e\to\infty} c'(e) = \infty$. The upshot is that effort is costly but helps generate both a) a private benefit (via, F, s and p) that would appeal to *Homo economicus*, and b) a part (via α and ρ) that would appeal to *Homo behavioralis*. Following DellaVigna and Pope (2018) and DellaVigna et al. (2016), we analyze the optimality conditions assuming two different functional forms for the cost of effort function : a power function and an exponential function i.e.,

$$c(e) = \frac{ke^{1+\gamma}}{1+\gamma}, \tag{2}$$

and

$$c(e) = \frac{kexp^{\gamma e}}{\gamma} \tag{3}$$

The power cost function (2) characterizes a constant elasticity of effort with respect to return to effort given by $1/\gamma$, while the exponential function (3) represents decreasing elasticity of effort with respect to return to effort given by $1/log(r/k)$, where r is the return to the effort. Workers' effort at different piece rates can be used to identify and structurally estimate both parameters of the cost-of-effort functions, namely, k and γ.

A worker solves the problem, $\max_{e_j \geq 0} U_j$. The interior solution is characterized by:

$$e_j^{\star} = c'^{-1}\left(s + \rho_j \mathbb{1}_{Gift} + \alpha_j v + p\right) \tag{4}$$

which, for the power cost function, yields :

$$e_j^{\star} = \left(\frac{s + \rho_j \mathbb{1}_{Gift} + \alpha_j v + p}{k}\right)^{1/\gamma},$$

and

$$e_j^{\star} = \frac{1}{\gamma}\, ln\left(\frac{s + \rho_j \mathbb{1}_{Gift} + \alpha_j v + p}{k}\right)$$

for the exponential form.

We start by making the simplifying assumption that workers are homogeneous given a treatment i.e., they will make the same effort choice as any other worker assigned to the same treatment. We later relax this assumption to account for heterogeneity in effort within a treatment. Our goal is to identify the parameters of the model just described. To that end, we design our treatments by varying the incentives and behavioral motivators for the workers.

2.1 Piece Rate Treatments

Here, all else same, each worker works on a task at a given piece rate of either 0, 3, 6 or 9 cents per unit of effort (calibrated to 100 points scored on the task). The piece rates generate income in addition to the \$1 fixed participation fee, F. By M-Turk standards, this amount of variation in piece rates is substantial enough to elicit significant changes in effort thereby allowing us to estimate the baseline parameters (s, k, and γ) which, in turn, are used to estimate other behavioral parameters.

Formally, in the piece rate treatments, a worker observes a piece rate p and then chooses effort e_j. There is no corresponding employer j present in these treatments. This shuts down altruism and reciprocity right away: for any worker, $\alpha_j = 0$ and $\mathbb{1}_{Gift} = 0$. The equilibrium efforts $e_j^\star$ in these treatments is thus given as:

$$e_p^\star = c'^{-1}(s+p) \text{ for } p \in \{0,3,6,9\}$$

The solution of effort has one behavioral unknown (s), and two unknowns from the cost function (k and γ). To back these out, we use effort corresponding to three different piece rates which gives us three equations to identify these parameters.

2.2 Altruism Treatments

In the altruism treatments, each worker is matched (see below for details) to an employer (truthfully) and he/she observes the (true) value of his/her effort to the matched employer. Specifically, each participant knows that an employer earns 1 cent for every 100 points scored by the matched worker. So as to not contaminate social preference with individual benefit, we set the piece rate to 0 in the three altruism treatments. In the first treatment (altruism baseline) a worker knows he/she has been matched to an employer but does not observe the employer's identity. In the 'altruism black' and 'altruism white' treatments, the worker observes the matched employer to be black and white, respectively.

Formally, in the altruism treatments, a worker observes the zero piece rate ($p = 0$), the value of the unit of effort to the employer j ($v = 0.01$), and then chooses effort e_j by maximizing (1). There is no gift from the employer implying $\mathbb{1}_{Gift} = 0$. The equilibrium efforts $e_j^\star$ in these treatments is, thus, given as:

$$e_j^\star = c'^{-1}\left(s + \alpha_j v\right) \text{ for } j \in \{Neutral, Black, White\}.$$

We are implicitly assuming that the altruism parameter can vary by the employer's group identity. For instance, $\alpha_{White} > \alpha_{Black}$ ($\alpha_{White} < \alpha_{Black}$) represents stronger (weaker) altruistic feelings for white as opposed to black employers. (As will be clear soon, all the workers in our sample are white which means α_j represents the strength of altruism a white worker feels for the jth employer.) Notice, since the piece rate is held fixed at 0 and reciprocity is shut out, the difference in effort provision between the 'altruism white' and 'altruism black' treatments is identifiable as resulting solely from the employer-race-dependent altruistic preferences of the workers. The three altruism treatments help us identify $\alpha_{Neutral}$, α_{Black}, and α_{White}, given the baseline parameters.

2.3 Altruism and Reciprocity Treatments

Altruism and reciprocity treatments (AR, henceforth) build on the altruism treatments and add a positive monetary gift (20 cents) from the employer to the worker. The remaining details are exactly the same as in altruism treatments. Thus, the equilibrium effort is given as;

$$e_j^{\star} = c'^{-1}\left(s + \alpha_j v + \rho_j\right) \text{ for } j \in \{Neutral, Black, White\}$$

As above, we are implicitly assuming that the reciprocity parameter may be different for each employer's group identity. In other words, controlling for the differences in altruism, the difference in effort between the treatments 'reciprocity white' and 'reciprocity black' is interpreted as resulting solely from the differential reciprocity preferences of the workers. The three AR treatments help us identify $\rho_{Neutral}$, ρ_{Black}, and ρ_{White} given the baseline and altruism parameters.

3 Experiment Design

The main goal of this study is to investigate the possibility of discrimination by workers towards their out-group employers in an online labor market. Our variable of choice is effort provision and the margin of choice is intensive. Our experiment is designed to ensure that observed differences in effort provision can only realize because of the race-dependent social preferences of workers. That is, if we detect any discrimination, it will be entirely driven by taste parameters; after all, we rule out the possibility of statistical discrimination by making it clear that worker choices in no way can affect their future earning prospects on M-Turk and employers will not get to make any payoff-relevant (or otherwise) choices after workers are done with the task.

3.1 Task

We need a task that is costly, effort-wise, to workers but is not meaningful in any way to a particular race. The task must require no special ability either. We settled on a button-pressing task as in DellaVigna and Pope (2018). The task involves alternating presses of "a" and "b" on a keyboard for 10 minutes. We chose it because it is simple to understand and has features that parallel clerical jobs: it involves repetition, it gets tiring (and boring), and therefore tests the motivation of the workers to stick to it and bring benefits to himself or his employer.

3.2 Race Revelation

We take the approach of revealing race via the revelation of skin color (Doleac & Stein, 2013). To that end, we record videos of employers in otherwise-identical scenarios as they read off a script explaining and demonstrating the task. The camera placement only captures the hand of the employer along with the movement of the fingers alternating 'a' and 'b' button presses. Other identifiers, such as the face, are not captured in the video to avoid psychological confounds often associated with faces, such as attractiveness and trustworthiness (Eckel & Petrie, 2011). The employer's hand is bare or covered (with full sleeves and latex gloves) depending on the assigned treatment. For black employers, we restrict the sample to participants with darker skin tone to avoid any ambiguity about the race of the person. We mute the voice for the videos in the neutral treatments. We program each video to play with subtitles to aid easier understanding of the instructions. The sample video links for each treatment are given in Table 1.

3.3 Experiment Flow

The experiment proceeds as follows: (1) First, we recruit employers, students from a major public university in the U.S. Midwest and record videos of them explaining the task, 2) next, we post a HIT on Amazon's Mechanical Turk inviting M-Turkers to take a screener survey, [8] (3) we invite those who meet the recruitment criteria (undisclosed) and consent to participate to initiate the experiment, (4) upon initiation, we assign each subject to one of the aforediscussed treatment groups. Following Czibor, Jimenez-Gomez, and List (2019), we use the blocked randomization design to assign subjects to treatments. We define blocks based on demographic information collected in the screener survey (Gender, Age, Race, Education, Income, Political Party Affiliation, and the Most-Lived U.S. state),[9] Next, (5) we present instructions to each subject in a pre-recorded video (based on the assigned treatment). We program our study to *require* each worker to watch the assigned video, (6) we elicit incentivised beliefs from each worker about their matched employer,[10,11] and 7) workers start to work on the task for a maximum of 10-minutes.

3.3.1 Piece Rate Treatments

In the piece rate (PR) treatments, each worker sees a video demonstrating a task with a script: "*On the next page, you will play a simple button-pressing task. The object of the task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible*

[8] Roughly 50% of M-Turkers are from the United States. Based on 2015 data, about 77% are non-Hispanic white and only 6% are non-Hispanic black (Hitlin, 2016). The results reported below are for U.S.-based white workers, by far the vast majority of workers on M-Turk and in our sample.

[9] See Cavaille (2018) for instructions on implementing sequential blocked randomization for online experiments.

[10] The elicitation of beliefs *before* workers start work on the task serves two purposes: 1) it provides us with data on workers' beliefs about the identity of their employer, and 2) it allows for the identity of the employer to become salient to the worker; importantly, it renders prominence to the seemingly-obvious fact that the worker is indeed matched to a real person whose payoff will be influenced by the worker's choices. We believe prior belief elicitation serves to increase salience of employer identity but does not amount to targeted priming courtesy the *between-subject* design of the study. It is important to note that workers' beliefs are elicited on a variety of identities (gender, age, income, and so on), not just race. We do not reveal our desire to know about their beliefs on race; it is just one of *six* different categories they are asked to report their beliefs on. As such, we are confident, our results are not tainted by experimenter demand effects. Parenthetically, in post experiment comments, not one worker identified ours to be a study about race or discrimination.

[11] To discourage random guessing in the belief elicitation part, participants are informed that an incorrect guess will lead to a deduction of 2 cents from their final earnings.

Table 1: Summary of treatments

Category (1)	**Treatment Wording** (2)	**Voice** (3)	**Skin Color** (4)	**Sample Video** (5)
Piece Rate (PR)	Your score will not affect your payment in any way.	Muted	Concealed	Link
	As a bonus, you will be paid an extra 3 cents for every 100 points that you score.	Muted	Concealed	Link
	As a bonus, you will be paid an extra 6 cents for every 100 points that you score.	Muted	Concealed	Link
	As a bonus, you will be paid an extra 9 cents for every 100 points that you score.	Muted	Concealed	Link
Altruism (A)	I will earn 1 cent for every 100 points that you score. Your score will not affect your payment in any way.	Muted	Concealed	Link
	I will earn 1 cent for every 100 points that you score. Your score will not affect your payment in any way.	Black	Black	Link
	I will earn 1 cent for every 100 points that you score. Your score will not affect your payment in any way.	White	White	Link
Altruism and Reciprocity (AR)	I will earn 1 cent for every 100 points that you score. In appreciation to you for performing this task, I have decided to pay you extra 20 cents as a bonus. Your score will not affect your payment in any way.	Muted	Concealed	Link
	I will earn 1 cent for every 100 points that you score. In appreciation to you for performing this task, I have decided to pay you extra 20 cents as a bonus. Your score will not affect your payment in any way.	Black	Black	Link
	I will earn 1 cent for every 100 points that you score. In appreciation to you for performing this task, I have decided to pay you extra 20 cents as a bonus. Your score will not affect your payment in any way.	White	White	Link

Notes: The table list all the treatments in this study. Each piece rate treatment differs just in the last line of the script, uses no audio, and conceals the skin color of the hand. Social preference treatments (altruism and reciprocity) begin with the introduction of the employer (in the first person), explain the task using the same script as in piece rate treatments and then differ only in the last paragraph of the script. Both 'Altruism' and 'Altruism and Reciprocity' categories have three treatments, each with black, white, and concealed skin tone of the employer (using gloves). In the social preference treatments of concealed skin tone, the ratio of black and white employers is 1:1.

for ten minutes. Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you alternate button pushes: just pressing the 'a' or 'b' button without alternating between the two will not result in points. Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used), or task will not be approved. Feel free to score as many points as you can." The final line is tailored to the assigned treatment (0, 3, 6 or 9 cents for PR-0, PR-3, PR-6 and PR-9 treatments respectively). The wording is provided in Table 1. Even though piece rates are framed in units of 100 points, workers are paid continuously for each point scored and are able to see the earned bonus in real time as they score points.

3.3.2 Social Preference Treatments

In the 'altruism' and 'altruism and reciprocity' treatments, each video starts with the introduction by the employer: "*Hi, I am another participant in this study who is matched to you. In this study, you will work on a simple button-pressing task, and I will earn some money depending on how well you do on the task.*" Thereafter, the script follows the same instructions as in piece rate treatments with the last paragraph being tailored to the social preference treatment in question. The wording is provided in Table 1. There are three treatments each in the category of 'altruism' and 'altruism and reciprocity'. Altruism-neutral (AN) and altruism and reciprocity-neutral (ARN) conceals the skin color of the employer in the video using latex gloves. The voice in the baseline treatments is also muted so as not to reveal any racial markers present in the voice. We recruit an equal number of black and white employers in the neutral treatments. The videos shown to workers in the altruism black (AB) (white (AW)), and altruism and reciprocity black (ARB) (white (ARW)) clearly reveal the black (white) skins of the employers, respectively.

While the 'altruism' and 'altruism and reciprocity' treatments give us a measure of the total effort, differencing out the respective baseline treatments will allow us to identify the pure altruism and reciprocity effects. In particular, the difference between AN treatment and PR-0 treatment gives us the race neutral altruism independent of race, while the difference between AB (AW) and PR-0 gives us workers' altruism towards the black (white) employer. Similarly, the difference between ARN treatment and AN treatment gives us the race neutral reciprocity independent of race, while the difference between ARB (ARW) and AB (AW) treatment gives us workers' reciprocity towards the black (white) employer.

3.4 Recruitment of Subjects

3.4.1 Recruitment of Employers

To recruit employers, we invite male student subjects over the age of 18 from a major public university in the U.S. Midwest who racially identify as either African American or Caucasian. We restrict our sample to male and U.S.-based employer-subjects to avoid confounds arising from identity effects of gender and nationality effects. Holding the sample size fixed and restricting it to one social identity give us extra statistical power and thereby ability to draw more credible inferences. We restrict the sample to employer subjects who are either black or white (we exclude Asians and Latinos, for example) because we believe our race-revelation mechanism works best in the context of these two races. We call these student subjects "employers" because they assign tasks to the workers who, in turn, work for these subjects and receive compensation (as is typical in most employer-worker relationships). Workers, at no point, know that the "employers" are students.[12] When an employer-subject arrived at the lab, they filled out a short demographic survey and was then randomly assigned to one of six social preference treatments. Based on the assigned treatment, subjects read from the script

[12]These employer subjects are framed as "other participant" to the workers so as not to introduce any imaginative effects from the use of make-believe language such as "employer".

and demonstrate the task on a video. Each subject was paid $5 for participation and an additional variable amount (average of $17.5) depending on their matched worker's performance. Our final sample include six employers in each social preference treatment (in all, 36 employers, 18 blacks and 18 whites).

3.4.2 Recruitment of Workers

We recruit U.S. based workers from Amazon's Mechanical Turk, a popular crowd-sourcing web-service that allows employers (called requester) to get tasks (called Human Intelligence Tasks (HITs)) executed by employees (called workers) in exchange for a wage (called reward). Mechanical Turk is a widely used platform for research in economics and allows access to a large pool of applicants at an affordable rate.[13]

We post a screener survey as the HIT on M-Turk with the following description "*Fill out this 2-minute screener survey to qualify for a study that starts immediately, take up to 15 minutes, and pays participation bonus $1 with scope to earn extra. You will be required to watch and listen to a video. Do NOT take this study on mobile.*". The responses to the screener survey allows us to pick participants that satisfy the criteria listed above. We allowed both black and white workers to participate. As per our pre-registration commitment, we recruited black workers only for race-salient, social-preference treatments. However, in the end, we could only recruit 711 (U.S.-based) black workers in the four social preference treatments combined. Power considerations, therefore, precluded their inclusion in the final analysis. Perforce, we restrict attention to white workers and study their effort choices for black versus white employers. We paid 15-cents to each potential subject for filling out the screener survey. On average, the workers in our sample earned $1.72 (including $0.15 for the screener survey, $1.0 for participation, and upto $0.1 for belief elicitation questions). These payments are sizable as per M-Turk standards for a 10-minute task and come close to the pro-rated,

[13] See Paolacci, Chandler, and Ipeirotis (2010) and Paolacci and Chandler (2014) for a discussion on demographic characteristics and representation of subjects from M-Turk.

federal minimum wage in the United States.

3.5 Pre-registration

We pre-registered the design on AEA RCT registry as AEARCTR-0003885. Since our task is the same as used in DellaVigna and Pope (2018), we can use results from their study to determine the sample size needed to achieve sufficient power for our study. DellaVigna and Pope (2018) find that the points scored across all treatments have a standard deviation of around 660. Assuming this standard deviation for each treatment and assuming a minimum detectable effect of 0.2 standard deviations between two treatments, we needed around 400 subjects in each treatment to have a power of 80 percent. This implies that we needed 400 x 10 = 4,000 observations in total for all ten treatments. We pre-registered the rule for sample size collection: we aimed to recruit 6,000 worker-subjects from M-Turk within the first three weeks of posting the experiment. Our data collection went slower than anticipated, and we ended up recruiting subjects from August 5th, 2019 to October 24th, 2019. In our registration, we had also planned to recruit self-identified black workers, which as explained above, did not work out.

4 Data

4.1 Employers

The demographic characteristics of the employer subjects in each treatment are presented in Table B2.

4.1.1 Pre-Testing of Videos

To verify whether the videos accurately reveal race, we test them using an independent sample of U.S.-based, white subjects from *Academic Prolific*, a data collection platform.

We used them instead of M-Turk to ensure that our M-Turk recruits could not have watched these videos before they participate in our experiment. Each subject was asked to identify the race of the person in a randomly-assigned video. See Figure A1 of the Online Appendix A for a graphical representation of average perception of race across treatments. Overall, race is correctly perceived more than 80 percent of the time for all the race-salient treatments – our race revelation mechanism works. For the race-neutral treatments, only less than 30 percent of the participants could guess the race, probably the result of random guessing. The pairwise comparisons of race perception among these treatments is presented in Table B3. The results suggest that the race-neutral treatments (AN and ARN) are statistically indistinguishable from each other and significantly different from race-salient treatments. The perception of race in the treatments AB and AW is statistically indistinguishable; however ARB is not perceived as accurately as ARW.

Participants also evaluated the videos in race-salient treatments for perception of skin color; the results are presented in Figure A2 of the Online Appendix A. Overall, blacks' skin is correctly perceived to be of darker tone and whites' of lighter tone. The pairwise comparisons of skin color perception among these treatments is presented in Table B3. The results suggest black treatments are statistically indistinguishable from each other and are significantly different from white treatments.

Finally, to check whether subjects in the videos were not perceived differently on soft personality traits such as friendliness, professionalism, clarity etc., we get these videos rated on those traits. The results for positive and negative traits are presented in Figure A3 and A4 of the Online Appendix A respectively. Pairwise comparisons of means across all the social-preference treatments suggest only the ARB treatment is perceived to be significantly higher on positive traits while all other treatments are statistically indistinguishable from each other on both positive and negative traits (see Table B4 of the Online Appendix B). This confirms that the only difference between the black and white employer' videos is the perceived race of the employer.

4.2 Workers

As per the pre-registration, we apply the following restrictions to the collected data: (1) we drop 17 participants who scored above 4,000 points as this is physically impossible in the 10- minute time-frame – likely, these users used some automated programs to score points;[14], 2) we drop 64 workers who scored zero points as this may reflect some malfunction or technical problem in the recording of points;[15] (3) we drop 4 observations of workers who participate more than once;[16] (4) we dropped two observations from workers who somehow managed to take this study from outside the United States.[17]

The final sample consists of 5,945 workers and the summary statistics are presented in Table B1 of the Online Appendix B. Our sample over represents women, young, educated, middle-income, and Democrats as compared to the U.S. labor force. This is typical of the population on online platforms. We present results of productivity by demographics in Table B9. Overall in our sample, men and younger workers are more productive than women and older workers respectively. We present test-of-balance of demographic variables across ten treatments in Table B5 of the the Online Appendix B. The treatments are balanced on all the observed variables, no surprise since we use blocked randomization to assign subjects to treatments. Since worker characteristics are balanced across treatments, there is no reason to believe that more/less productive workers are assigned to any specific treatment.

[14]We instructed each worker up-front to not use any automated scripts/programs .

[15]These workers are spread across all treatments, and there is no systematic difference in workers scoring zero points for any particular treatment or employer.

[16]A worker can participate in our study only once; these exceptions must be an error on the part of M-Turk.

[17]The study was restricted to U.S.-based workers. Presumably, these participants used a proxy server or VPN to mask their origin but we could spot them from the GPS coordinates recorded by Qualtrics.

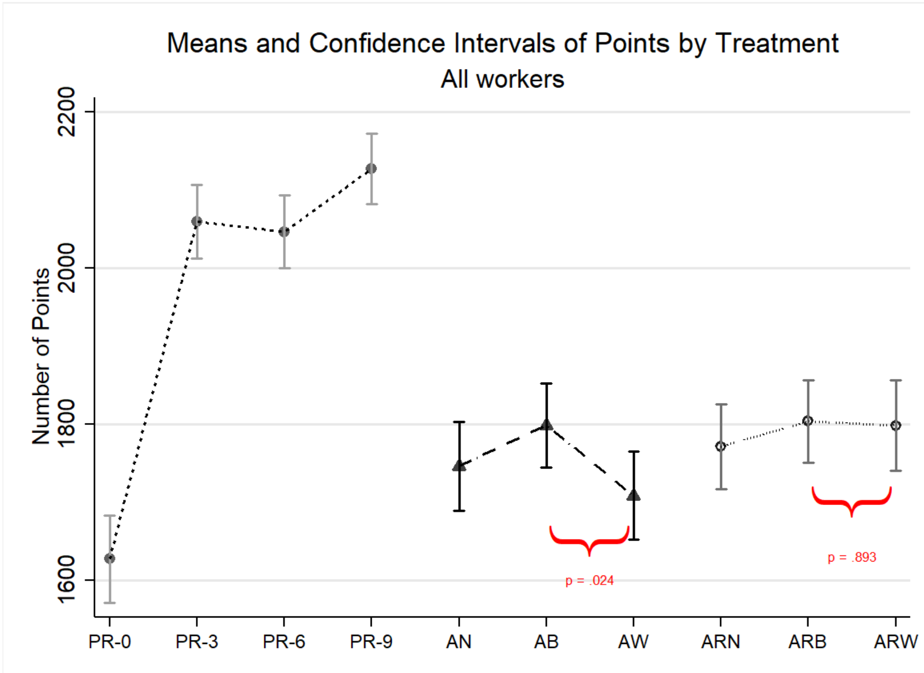

Figure 1: Effort by Treatment - All Workers

Notes: This figure presents the average score and confidence interval for each of ten treatments for all workers. Each treatment has about 590 participants.

5 Results

We present average effort by workers (recall, all our workers are white) against each treatment in column 1 of Table 2 and in Figure 1. Overall, it is evident that incentives have a strong effect on effort, raising performance from 1627 points (PR-0) to 2060 points (PR-3) and 2127 points (PR-9).[18] However, the average effort for 3-cent and 6-cent treatments is statistically the same, reflecting a low elasticity of effort beyond an initial increase in effort from 0 to 3 cents. The standard error for the mean effort per treatment is around 30 points or less, implying that differences across treatments larger than 85 points are statistically significant.

As discussed in Section 3.3.2, we define *altruism* as difference between the altruism

[18] Workers' positive effort in the 0-cent treatment is explained by the parameter s of the model in Section 2, which possibly captures the intrinsic motivation of the worker or her sense of duty/fairness or even even her unsubstantiated fear of being rejected for not scoring enough points.

Table 2: Effort by Treatment

	(1) All Workers		(2) Correctly Perceived Race	
	N	Mean (s.e)	N	Mean (s.e)
PR-0	599	1627.07	599	1627.07
		(28.56)		(28.56)
PR-3	595	2059.83	595	2059.83
		(24.19)		(24.19)
PR-6	592	2046.68	592	2046.68
		(23.62)		(23.62)
PR-9	588	2127.37	588	2127.37
		(23.01)		(23.01)
AN	591	1746.06	261	1724.87
		(29.15)		(43.70)
AB	601	1798.37	494	1807.68
		(27.55)		(29.58)
AW	592	1708.09	557	1715.24
		(28.90)		(29.52)
ARN	608	1771.15	265	1766.99
		(27.95)		(41.63)
ARB	590	1803.61	470	1818.78
		(26.95)		(29.73)
ARW	589	1798.23	561	1803.75
		(29.58)		(30.33)
Total	5945	1848.08	4982	1865.98
		(8.80)		(9.49)

Notes: The table presents the effort choices in each treatment. Column 1 reports the effort choices by all the workers, column 2 reports the effort choices by workers who were able to correctly perceive the race of the employer as neutral, black or white in social preference treatments.

treatments and the piece rate 0-cent treatment and *reciprocity* as the difference between the 'altruism and reciprocity' treatments and the respective 'altruism' treatments. How do we detect altruism? We compute the average effort of workers in the AN treatment; recall, these are workers who know they are matched to an employer but don't know his race. We compare that to the average effort of workers who are not matched to any employer and are offered a 0-cent piece rate (PR-0). We find statistically significant evidence for altruism: workers put more effort in the AN treatment as compared to the PR-0 treatment. The one cent return to the employer induces an effort of 1746 points as compared to 1627 points in the PR-0 treatment. This difference of 119 points is significantly different from 0 and is presented in Figure 2.

Next, we compare the strength of the altruism preference across black and white employers. To that end, we compute the average effort of workers in the AB treatment to those in the AW treatment. Strikingly, workers are significantly more altruistic towards black employers than white employers. The average effort for black employers is 1798 points, which is significantly higher ($p = 0.024$) than the same for white employers (1708 points). However, note that in the altruism treatments, average effort for any race employer is not significantly different from the same for a race-neutral employer.These differences are more clearly presented in Figure 2.

In the 'altruism and reciprocity' treatments, the worker receives an unanticipated gift of 20-cents (in addition to all other earnings) from the employer, unconditional on performance. This gift in the ARN treatment does not induce a significant increase in effort as compared to the AN treatment (1771 points in the ARN treatment as compared to 1746 points in the AN treatment). This result is consistent with the literature which finds weak evidence for positive reciprocity (such as Kube, Maréchal, and Puppe (2006)). However, as shown in the Figure 2,the reciprocity, derived by taking a difference between the 'altruism and reciprocity' treatments and respective altruism treatments, varies by employer's race. The reciprocity towards the neutral and black

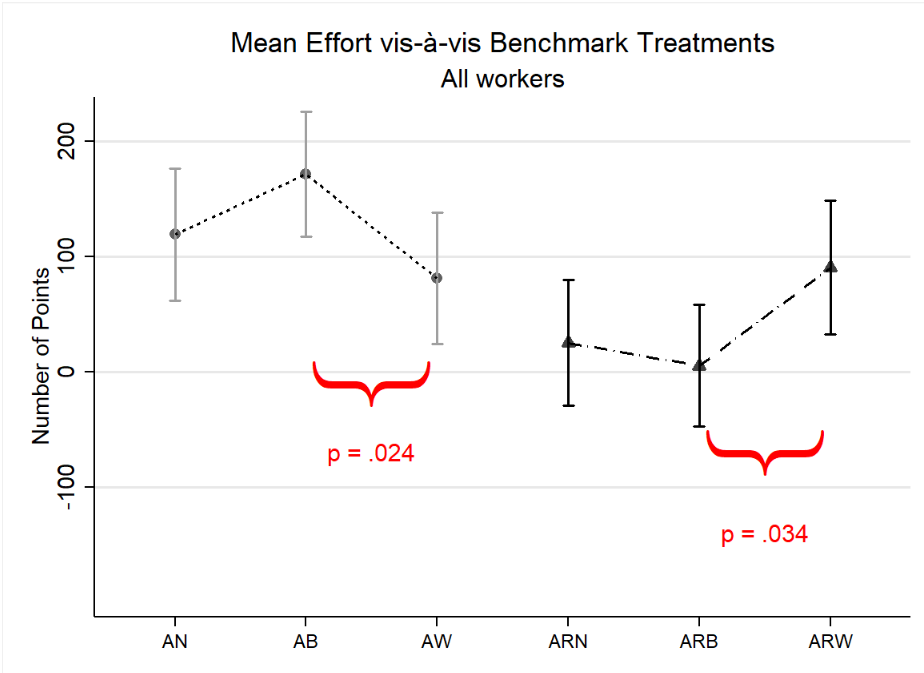

Figure 2: Net effect of social preferences - All Workers

Notes: This figure presents the average net effect and confidence interval for each of the social preference treatments for all workers. Net effect for altruism (reciprocity) is calculated by taking an average of deviation of points scored in each altruism (altruism and reciprocity) treatment from mean of points scored in 'Piece Rate - 0 cent' (respective altruism) treatment. Each treatment has about 590 participants.

employer is statistically zero, while that towards the white employer is positive and statistically significant. Turns out the difference between the reciprocity towards the white and black employer is positive and statistically significant as well (p=0.034). This indicates that, on average, workers reciprocate towards the white employers but not towards the black employers. In fact, the reciprocal response towards white employers completely counter the positive altruistic response towards the black employers, so overall, black and white employers receive similar effort from the workers, as is clear from Figure 1.The altruism and reciprocity responses in each treatment for workers who correctly guessed the race of the employer is given in the Figure B8 of Online Appendix B.

Although our treatments are balanced on observed worker and employer characteristics, for robustness sake we present the regression results from regressing "Points" scored on the employer racial identity, with race neutrality as the omitted category, and controlling for these variables in the Table 3. We observe that the difference in altruism between the race salient treatments and race neutral treatment continues to be insignificant. Further, workers' pro-altruistic response for black employers stays significantly different from white employers, even after controlling for the demographic variables and the perception about the employer's income, age, and education. This is indicated by 'Black-White' reported at the bottom of Table 3 and highlight the fact that the higher altruism towards blacks is not driven by the differences in beliefs about the employer's income, age, and education. Similarly, the difference in pure reciprocity between race salient treatments and race neutral treatment is largely insignificant. However, workers are significantly more reciprocal to the 20 cents gift towards the white employers relative to the black employers, even after controlling for the demographic variables and the perceptions about employer's income, age and education. This discrimination in reciprocity against the black is evident from 'Black-White' at the bottom of Table 3.

5.1 Distribution of Effort

Beyond average effort, we present the distribution of effort from all the treatments in Figure A5 of the Online Appendix A and by each treatment in Figure A6 of the Online Appendix A. Overall, very few workers score below 500 points and even fewer score above 3000 points.

Figure A7a of the Online Appendix A presents the cumulative distribution function for the piece rate treatments. Incentives induce a clear rightward shift in effort relative to the PR-0 treatment. However, there is not much change in effort between the PR-3 and the PR-6 treatments. Figure A7b of the Online Appendix A shows strong

Table 3: Social Preference Treatments - Robustness

	Altruism			Reciprocity		
	(1)	(2)	(3)	(4)	(5)	(6)
Black or African American	52.31	53.68	56.49	-19.85	-15.16	-38.71
	(40.31)	(40.74)	(41.42)	(39.74)	(40.43)	(40.92)
White or Caucasian	-37.97	-30.84	-33.82	65.06	72.30*	39.50
	(40.46)	(41.07)	(41.78)	(39.76)	(40.67)	(41.60)
Constant	119.0***	21.78	78.86	25.09	55.54	35.70
	(28.62)	(254.7)	(262.6)	(27.89)	(267.4)	(278.1)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Employer Perception	No	No	Yes	No	No	Yes
N	1784	1706	1701	1787	1719	1717
Black - White	90.28**	84.53**	90.30**	-84.91**	-87.46**	-78.21*
	(40.29)	(40.89)	(41.17)	(40.06)	(40.94)	(41.20)

Notes: The table presents, in column 1, 2, and 3, the estimates from an OLS regression of deviation of points scored in the altruism treatments from the mean of PR-0 treatment and, in column 4, 5, and 6, the deviation of points scored in the 'altruism and reciprocity' treatments from the mean of respective altruism treatment on the employer's race. The omitted category is the employer with concealed race. Demographic controls include age, gender, education, income, political affiliation and the voting pattern of the most lived state (red, blue, or swing) of the worker. Employer Perception include worker's belief about the income, age, and education of the employer. Black - White represents the difference in the coefficients of black and white employers in each model. Standard errors in parentheses. for p < 0.10, ** for p < 0.05, and *** for p< 0.01.*

evidence for altruistic preferences as observed by the clear rightward shift of the effort distribution in the altruism treatment as compared to the PR-0 treatment (ranksum test, p-value = 0.002). The effort distribution in the ARN treatment is indistinguishable from the AN treatment, implying a lack of reciprocal preferences (ranksum test, p-value = 0.57). Figure A7c of the Online Appendix A plots pure race salient altruism and reciprocity, and it shows that altruism is stronger towards blacks as compared to whites,while the cumulative density function for reciprocity-black is dominated by reciprocity-white treatment, indicating significantly lower reciprocal response towards the blacks. Quantile regression estimates for effort (Table B7 of Online Appendix B) show that black employers get higher effort than white employers at both the 0.25 and 0.5 quantile for the altruism treatments. This shows that the altruistic response for the black employers is mainly coming from the *lower part of the* effort distribution. On the other hand, the discrimination in reciprocity against the black employers arises mainly at the 0.5 and 0.75 quantile of the effort distribution. The Kolmogorov-Smirnov test of equality of distribution functions is presented in Table B6 of Online Appendix B.

5.2 Evolution of Effort

We present the evolution of effort over the 10-minute period in Figure A8 of Online Appendix A. Figure A8b and A8c shows that, in the social preference treatments, effort declines over time presumably due to boredom and tiredness. And yet, interestingly, the PR treatments are able to sustain consistently higher effort throughout the entire time interval (Figure A8a), with workers in the PR-9 treatment pushing extra hard near the end. In the altruism treatments (Figure A8b), the effort for the black employers (AB) is consistently higher than the effort for the white employers (AW) throughout the 10-minutes period. On the other hand, in reciprocity (Figure A8c), the effort for the black employer is consistently lower than the effort for the white employers, indicating that the average discrimination in reciprocity against the blacks discussed above, is borne out throughout the entire 10 minute period. .

5.3 Heterogeneity

5.3.1 Heterogeneity by Demographics

To examine the heterogeneity in our average treatment effects based on demographic characteristics of the sample, we present the differences in treatment effects in Table 4 for both altruism and reciprocity. Overall, we do not find evidence of heterogeneity in treatment effects on the basis of gender, age, and education in both altruism and reciprocity. However, we do find some evidence of heterogeneity in reciprocity. We find that higher income workers reciprocate significantly more to the black employers (vis-à-vis white employers) as compared to lower income workers.Needless to say that our tests for heterogeneity in treatment effects are possibly under-powered and as such it is hard to precisely identify a null effect.

5.3.2 Heterogeneity by the share of black population in the neighborhood

Following Andreoni, Payne, Smith, and Karp (2016), we explore the effects of local racial composition on social preferences of the workers in our sample. We condition on the zip code level racial composition of the worker and examine the difference in effort provided to black versus white employers. Table 5 presents the conditional average treatment effects for top and bottom quantile of the share of black population, for workers who correctly perceived the employer race for both altruism and reciprocity treatments. Consistent with the hypothesis 'familiarity breeds contempt', we find that workers from neighborhoods with lower share of black population are significantly more altruistic towards black employers. We do not find evidence for differential reciprocal response on the basis of share of black population in the neighborhood.

Table 4: Heterogeneity by Demographics

	(1) Altruism	(2) Reciprocity
Employer Race		
White	-113.6	77.49
	(112.0)	(111.9)
Gender		
Female	-151.8***	-142.5**
	(57.40)	(57.75)
White × Female	99.12	93.07
	(81.74)	(82.53)
Age		
35 and above	-41.70	-142.9**
	(57.40)	(57.01)
White × 35 and above	-19.01	5.892
	(81.87)	(81.27)
Education		
College and above	-109.0*	-160.7***
	(60.53)	(61.04)
White × College and above	-110.8	94.07
	(85.69)	(86.46)
Income		
$\geq$ \$45,000	65.78	153.9**
	(59.30)	(60.01)
White × $\geq$ \$45,000	30.76	-258.9***
	(84.40)	(85.27)
Party Affiliation		
Democrat	-85.17	-107.1
	(66.99)	(67.38)
Republican	33.12	-40.87
	(74.23)	(74.28)
White × Democrat	139.0	115.0
	(95.84)	(95.38)
White × Republican	-16.87	116.8
	(104.8)	(107.2)
State Voting Pattern		
Blue	-82.25	4.238
	(64.43)	(65.03)
Red	74.99	2.641
	(76.72)	(76.32)
White × Blue	23.33	-99.36
	(92.51)	(92.78)
White × Red	-174.9	-31.83
	(107.2)	(109.3)
Constant	347.3***	212.3***
	(76.18)	(77.00)
Observations	1171	1149

*Notes: The table presents the differences in average treatment effects by the demographics of the workers for both altruism (column 1) and reciprocity (column 2). The omitted employer is the Black employer. The reference categories for gender, age, education, income, party affiliation, and state voting pattern are female, below 35, below college, under \$45,000, independent, and swing state, respectively. Standard errors in parentheses. * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.*

Table 5: CATE by Black Share Quantiles - Bottom and Top

	Altruism		Reciprocity	
	(1)	(2)	(3)	(4)
	Lower Share	Higher Share	Lower Share	Higher Share
White	-184.0***	-57.15	44.62	75.02
	(65.09)	(59.62)	(62.96)	(64.87)
Constant	227.6	-195.6	-303.9	-127.3
	(532.3)	(390.8)	(543.6)	(387.5)
Demographic Controls	Yes	Yes	Yes	Yes
Employer Perception	Yes	Yes	Yes	Yes
Observations	487	515	498	477

Notes: The table presents the conditional average treatment effect by the bottom and top quantile of the share of black population in the worker's zip code for both altruism (column 1 and 2) and reciprocity (column 3 and 4). Measure of conditional treatment effect is obtained by restricting to workers who could correctly perceive the employer race and running an OLS regression of deviation of points scored in altruism treatments from mean of PR-0 treatment (column 1 and 2) and deviation of points scored in 'altruism and reciprocity' treatments from the mean of respective altruism treatments (column 3 and 4) on Employer Race for bottom (column 1 and 3) and top (column 2 and 4) quantile of share of black population in the worker's zip code while controlling for demographics and employer perception. Standard errors in parenthesis. for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.*

5.3.3 Heterogeneity by Geographical Area

It is a well established fact that racial disparities are not equally distributed across the U.S. We present the summary of workers performance by their geographical area in Table 6. Interestingly, there is an evidence in favor of workers from South being relatively more altruistic to black employers. This is surprising given that the average implicit bias against blacks (see next subsection) in the South is higher than in the other regions of the U.S. As for reciprocity, workers provide higher effort for white employers relative to black in all the geographical areas, however, not significantly so.

5.3.4 Heterogeneity by Implicit Biases

We examine the heterogeneity in treatment effects based on the implicit biases of workers as measured by the implicit association test (IAT). IATs are widely used in social psychology to measure implicit and unconscious biases towards a particular group. The test involves categorizing two sets of words to the left or right hand side of the computer screen. The implicit bias is measured by a time difference in associating good or bad words to the relevant group identities. The idea is that making a response is easier when closely related items share categorization to the same side of the screen. In case of race IAT, we would say that one has an implicit preference for white people relative to black people if they are faster to categorize words when white face and good words (friend, glorious, enjoy, joyous, terrific, beautiful, magnificent, and fabulous) share a response key and black faces and bad words (detest, poison, nasty, disgust, pain, despise, sadness, evil) share a response key, relative to the reverse.

For this study, we did not conduct IAT test for individual workers. Nor did we employ survey measures so as to avoid revealing the purpose of the study.[19] Instead, we proxy the IAT score of individual worker by using the geo-coded race IAT data

[19] M-Turkers often communicate with each other on various platforms; as such, we wanted to make sure that the purpose of the study is not broadcasted even after the worker is done with the task.

Table 6: Heterogeneity by Geographical Area

(a)

	Altruism			
	(1)	(2)	(3)	(4)
	North East	Mid West	South	West
White or Caucasian	16.42	-76.81	-155.4**	18.00
	(96.74)	(90.16)	(67.89)	(111.3)
Constant	717.6**	-140.1	-175.1	150.8
	(314.5)	(289.2)	(432.0)	(621.0)
Demographic Controls	Yes	Yes	Yes	Yes
Employer Perception	Yes	Yes	Yes	Yes
Observations	190	243	409	206

(b)

	Reciprocity			
	(1)	(2)	(3)	(4)
	North East	Mid West	South	West
White or Caucasian	14.27	98.22	55.48	28.06
	(111.4)	(77.90)	(75.22)	(112.5)
Constant	1506.7*	-562.2	-857.3	-147.1
	(874.7)	(449.8)	(758.1)	(822.7)
Demographic Controls	Yes	Yes	Yes	Yes
Employer Perception	Yes	Yes	Yes	Yes
Observations	186	286	360	198

Notes: The table presents the conditional average treatment effect by the geographical location of the worker for altruism (table (a)) and reciprocity (table (b)). Measure of conditional treatment effect is obtained by restricting to workers who could correctly perceive the employer race and running a regression of deviation of points scored in altruism treatments from mean of PR-0 treatment (panel a) and deviation of points scored in 'altruism and reciprocity' treatments from the mean of respective altruism treatment (panel b) on Employer Race for each geographical region while controlling for demographic and employer perception. Standard errors in parenthesis. for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.*

by Project Implicit, which provides historical record of tests taken on the project's website. These tests can be taken by anyone from anywhere in the world. For our purpose, we restrict to white individuals from the United States and use the data from more than two million test takers between 2006 to 2018. We map the county level (lowest available resolution) IAT score to workers in our sample based on the worker's geographic location. Our worker sample comes from 190 counties spanning all 50 states in the U.S.

Typical thresholds found in the literature (Greenwald, Nosek, & Banaji, 2003; Hahn, Judd, Hirsh, & Blair, 2014; Rooth, 2010) are as follows: IAT scores below -0.15 indicate some preference for minorities; scores between -0.15 and 0.15 indicate little to no bias; scores between 0.15 and 0.35 indicate a slight bias against minorities; and scores above 0.35 show moderate to severe bias against minorities. The average score (standard deviation) of white test takers, as proxied by the county level average IAT score, in our sample is 0.38 (0.42) implying, on average, white people have moderate to severe implicit bias against blacks. Like black share, we explore the effects of local IAT score on the social preferences of workers in our sample. We condition on the county level IAT score of the worker and examine the difference in effort provided for black versus white employers. Restricting to two quantiles of IAT score clearly shows (Table 7) that workers with lower implicit bias are significantly more altruistic towards black employers vis-à-vis white employers. On the other hand, those with higher implicit bias are significantly more reciprocal towards white employers relative to the black employers. In other words, the discrimination in reciprocity, which we observe towards black employers, emanate from workers who live in counties with relatively high implicit bias against the blacks.

Table 7: CATE by IAT Quantiles - Bottom and Top

	Altruism		Reciprocity	
	(1)	(2)	(3)	(4)
	Lower Bias	Higher Bias	Lower Bias	Higher Bias
White or Caucasian	-172.4***	-25.33	-0.576	135.5**
	(64.51)	(58.33)	(63.76)	(60.10)
Constant	-408.4	537.3	-341.1	208.3
	(467.2)	(413.2)	(408.2)	(503.1)
Demographic Controls	Yes	Yes	Yes	Yes
Employer Perception	Yes	Yes	Yes	Yes
Observations	529	513	526	495

Notes: The table presents the conditional average treatment effect by the bottom and top quantile of the IAT score of the worker's county for both altruism (column 1 and 2) and reciprocity (column 3 and 4) treatments. Measure of conditional treatment effect is obtained by restricting to workers who could correctly perceive the employer race and running a regression of deviation of points scored in altruism treatments from mean of PR-0 treatment (column 1 and 2) and deviation of points scored in the 'altruism and reciprocity' treatments from the mean of respective altruism treatment (column 3 and 4) on Employer Race for bottom (column 1 and 3) and top (column 2 and 4) quantile of IAT score while controlling for demographics and employer perception. Standard errors in parenthesis. for p < 0.10, ** for p < 0.05, and *** for p< 0.01.*

6 Estimates of Behavioral Parameters

We designed our experiment to go with the structural model outlined in Section 2. The advantage of designing field experiments on the basis of a model of behavior is that it allows researchers to estimate the nuisance parameters in the environment that are relevant to decision making (DellaVigna, 2018). Because of the simplicity of our task, there are only three nuisance parameters we need to estimate. We use data from the piece rate treatments to identify these parameters. Subsequently, we estimate the deeper behavioral parameters of interest using data from the social preference treatments. We closely follow the estimation procedure in DellaVigna and Pope (2018)
.

6.1 Minimum-Distance Estimation

We first use minimum-distance estimation method to estimate these parameters. In minimum distance estimation, one identifies the set of moments in the data (average effort) and then finds the set of model parameters that minimizes the distance between the empirical moments and the theory-predicted moments. To estimate nuisance parameters, we use the average effort corresponding to the three piece rates (0 cents, 3 cents and 9 cents), to estimate γ, s, and k. Specifically, in the case of the power cost function, to estimate nuisance parameters, we use first moments from the piece rate treatments and solve the following equations

$$\bar{e}_p = \frac{1}{\gamma}\left[log(s+p) - log(k)\right] \text{ for } p \in \{0, 0.03, 0.09\}$$

where $\bar{e}_p$ is the average effort in the piece rate p treatment. These parameters estimates are used to draw the marginal cost and marginal benefit curve in Figure 3.

Once these parameters are estimated, we use average effort corresponding to AN, AB and AW treatment to estimate behavioral parameters $\alpha_{Neutral}$, α_{Black}, and α_{White}

respectively. Specifically, for the power cost function, we solve the following equations for α_j for $j \in \{Neutral, Black, White\}$ taking estimates of γ, s, and k as given

$$log(\bar{e}_{\alpha_j}) = \frac{1}{\gamma}[log(s + \alpha_j v) - log(k)] \text{ for } j \in \{Neutral, Black, White\}$$

where $\bar{e}_{\alpha_j}$ is the average effort in the altruism j treatment.

Similarly, to calculate reciprocity parameters for neutral ($\rho_{Neutral}$), black (ρ_{Black}) and white (ρ_{White}) employers, we use average effort from ARN, ARB, and ARW treatments and solve the following equations taking estimates of γ, s, k, and α_j for $j \in \{Neutral, Black, White\}$ as given:

$$log(\bar{e}_{\rho_j}) = \frac{1}{\gamma}[log(s + \rho_j + \alpha_j v) - log(k)] \text{ for } j \in \{Neutral, Black, White\}$$

where $\bar{e}_{\rho_j}$ is the average effort in the reciprocity j treatment.

Estimates using the exponential cost function are similarly calculated. Table 8 presents the parameter estimates for power cost function (column 1) and exponential cost function (column 3). The standard errors for these parameter estimates are estimated using a bootstrap procedure with a thousand draws.

6.2 Non-Linear Least Squares Estimation

The minimum distance estimator solely relies on the moment, and hence, does not use all the variation in the data. There are methods such as maximum likelihood and non-linear least squares that can be used to estimate these parameters using all the variation present in the data. We use non-linear least square method to estimate these parameters allowing for the heterogeneous cost of effort. Allowing for a heterogeneous marginal cost of effort in 1, we assume for a worker i, for a power cost case, $c(e_{ij}) = \frac{ke_{ij}^{1+\gamma}}{1+\gamma} exp(-\gamma\epsilon_{ij})$ with $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. The first order condition 4 can then be written as;

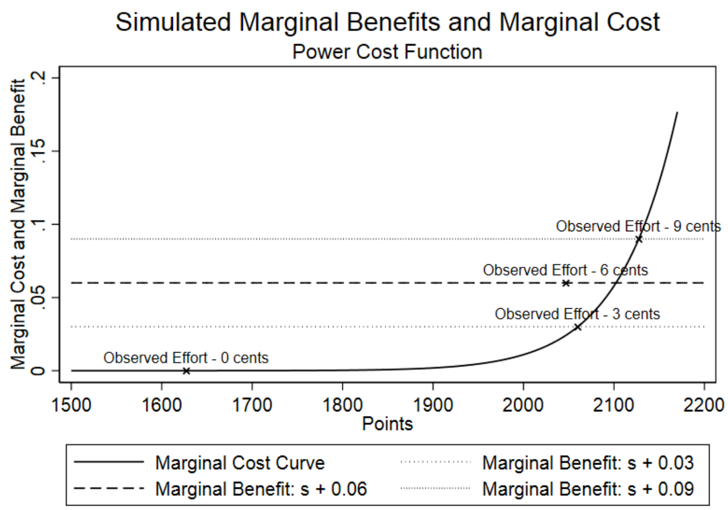

Figure 3: Illustration of the Model: Marginal Benefits and Cost Curves
Notes: The figure presents the marginal benefit and marginal cost curves using minimum-distance estimates for the power cost function.

$$s + 1_{Gift}\rho_j + \alpha_j v + p - ke_{ij}^{\gamma} exp(-\gamma\epsilon_{ij}) = 0$$

Taking the last term to the right and taking logs, we obtain

$$log(s + 1_{Gift}\rho_j + \alpha_j v + p) + \epsilon_{ij} = log(k) + \gamma log(e_{ij}) - \gamma\epsilon_{ij}$$

Solving for $log(e_{ij})$, we obtain the estimating equation

$$log(e_{ij}) = \frac{1}{\gamma}[log(s + 1_{Gift}\rho_j + \alpha_j v + p) - log(k)] + \epsilon_{ij}. \tag{5}$$

Similarly using exponential cost function, we get

$$e_{ij} = \frac{1}{\gamma}[log(s + 1_{Gift}\rho_j + \alpha_j v + p) - log(k)] + \epsilon_{ij}. \tag{6}$$

Equations 5 and 6 can be estimated using non-linear least squares (NLS). Table 8 presents the NLS parameter estimates for power cost function (column 2) and exponential cost function (column 4). The NLS parameter estimates are nearly identical to those computed with minimum-distance estimation for the exponential cost case. The model predictions are also very similar .

The NLS estimates for the power cost function yield a lower curvature than the minimum-distance estimates ($\hat{\gamma}_{NLS} = 20.29$ versus $\hat{\gamma}_{MD} = 34.05$). The NLS model matches expected log effort, while the minimum-distance matches the log of expected effort. Both NLS and minimum-distance fit the in-sample moments and make similar predictions for the 6-cent piece rate treatment.

The parameter estimate for 'altruism black' is significantly higher than 'altruism white' in all the specifications, indicating that white workers have significantly higher altruistic preferences for black employers as compared to white employers (p-value$<$0.01). Regarding reciprocity, the choice data indicates that the reciprocity parameter towards the black is significantly lower than that towards the white (p-value$<$0.01). Even though the parameter values are close to zero, but they translate to meaningful differences in effort provided to black and white employers at the zero piece rate. Figure 4 presents the simulated effort for neutral, black and white employer using parameter estimates along with zero social-preference case. Black employers receive around five percent higher effort than white employers at the zero piece rate owing to a worker's altruistic preference. The reciprocity parameters are small as compared to altruism parameter, so the simulated effort for reciprocity are indistinguishable for black and white employers. The difference between black and white employers becomes negligible at higher piece rates because workers respond much more to monetary incentives as compared to social preferences.

Table 8: Parameter Estimates

	Power Cost		Exponential Cost	
	Minimum Distance	NLS	Minimum Distance	NLS
	(1)	(2)	(3)	(4)
Baseline Parameters				
γ	34.05	20.30	0.0163	0.0163
	(7.3)	(8.85)	(.0102)	(.00807)
s	0.00000977	0.00000802	0.0000264	0.0000264
	(.000199)	(.000032)	(.000327)	(.000101)
k	4.50e-115	2.98e-70	8.58e-17	8.58e-17
	(3.6e-49)	(2.5e-68)	(3.1e-09)	(1.5e-15)
Altruism Parameters				
α_{Neutral}	0.00983	0.000426	0.0156	0.0156
	(.00739)	(.0017)	(.00966)	(.0427)
α_{Black}	0.0285	0.000776	0.0402	0.0402
	(.0181)	(.00274)	(.0214)	(.0953)
α_{White}	0.00413	0.000270	0.00722	0.00722
	(.00367)	(.00129)	(.00531)	(.0215)
Reciprocity Parameters				
ρ_{Neutral}	0.0000676	0.0000272	0.0000921	0.00124
	(.000134)	(.000103)	(.000182)	(.00318)
ρ_{Black}	0.0000307	0.0000395	0.0000381	0.00220
	(.000271)	(.00014)	(.000321)	(.00513)
ρ_{White}	0.000243	0.0000255	0.000328	0.00200
	(.000216)	(.0001)	(.00027)	(.00477)
Implied effort - 6-cents	2102	7.746	2102	2102.4

Notes: This table reports the structural estimates of the model in section 2. Column (1) and (3) use a minimum-distance estimator employing three moments (average effort in three piece rate treatments) and three parameters (γ , s and k), and is thus exactly identified. Column (2) and (4) use a non-linear least squares employing individual effort in all the treatments and thus estimating all the parameters simultaneously. We use power cost (column 1 and 2) and exponential cost (column 3 and 4) function to estimate the model. Implied effort is calculated using estimated parameters for each model. The observed effort for 6-cents treatment is 2047 points or log 7.624. For the altruism parameters, the baseline parameters are taken as given and the average effort for neutral, black, and white employers is used to estimate $\alpha_{Neutral}$, α_{Black}, and α_{White} from the altruism treatments. Similarly for the reciprocity parameters, the baseline and altruism parameters are taken as given and the average effort corresponding to reciprocity neutral, reciprocity black, and reciprocity white is used to estimate $\rho_{Neutral}$, ρ_{Black}, and ρ_{White}. Standard errors for minimum-distance estimator are calculated by taking a bootstrap sample of 1000 draws and recalculating these parameters for each draw.

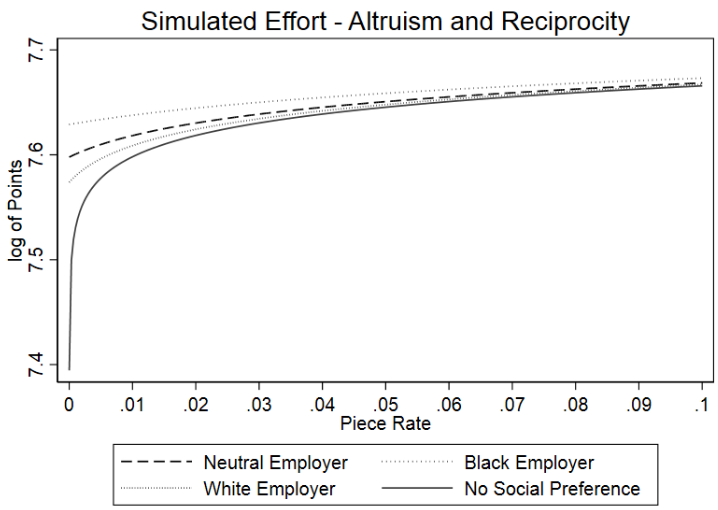

Figure 4: Simulated Effort by Employer Race at Different Piece Rates

Notes: The figure presents the simulated effort using the parameter estimates from table 8 for power cost, minimum distance specification. Neutral/Black/White employer uses the respected social preference parameter estimates to calculate the predicted effort at each piece rate. No Social Preference assumes that altruism and reciprocity estimates are zero.

7 Conclusion

Economic historians record a time in U.S. labor history when white workers openly militated against receiving orders from (or working under) black supervisors. While overt racism was implicated in the past, it is behavioral differences that lie at the root of racial inequality in contemporary America (Loury, 1998). What are these behavioral differences? Now that overt racism is either illegal or difficult to practice openly, have white workers stopped discriminating against black employers? This paper uses insights from behavioral and experimental economics to shed a bit of light on this enduring issue in American labor markets. The narrower question we ask is, do white workers, given a considerable degree of discretion over work effort, display discrimination in race-dependent, social preferences toward their black employers?

The experimental setting is an online labor market - Amazon's Mechanical Turk (M-Turk). In this online economy, workers and employers are at arms-length, and the worker is involved in a real-effort task for a pre-assigned, non-fictitious, black, or white employer. The possibility of race-dependent social preferences is activated by unobtrusively revealing the employer's race to the matched worker. In such a setting, we detect statistically significant evidence for altruism: white workers put significantly more effort when they know their work benefits their employer as compared to one where neither the worker nor the employer benefits from worker effort. Contrary to the historical narrative, white workers show more altruism toward black employers than white ones. Not only is this finding statistically significant at the 2% level, but the difference in effort provision is also economically big as well. This is important and rare evidence for prosociality. Strikingly, though, white workers display significantly higher reciprocity for white employers. To our knowledge, this constitutes the first evidence of worker-to-employer discrimination. Interestingly, when the altruism and reciprocity responses are taken in tandem, the effect of the prosociality towards blacks gets thoroughly *washed out* by the discrimination in reciprocity against them. This means, overall, we do not detect differential, race-based effort provision of white workers toward black bosses.

Our work finds compelling evidence of a racially-motivated breakdown of reciprocal links that form the bedrock of labor market exchange in any economy. Specifically, in our sample, white workers appear lukewarm in their effort response to a gift from a black employer. In contrast, the same from a white employer evokes a significant, positive response. This discrimination in reciprocity implies white employers are 'rewarded' more for the same gift by white workers as compared to their black counterparts. It is not a big stretch from there to see that if what we detect is wide-spread, white bosses will be more successful in leadership positions than black ones. Indirectly, this may also be relevant for explaining the diminished presence of blacks in leadership positions.

How do we view our results and their value-added in the context of the broader literature surveyed in Bertrand and Duflo (2017)? Starting from Bertrand and Mullainathan (2004), study after study, "close replications" of the Bertrand and Mullainathan (2004) correspondence-study approach, finds overwhelming evidence of discrimination in hiring decisions by employers against workers from racial and ethnic minorities. Bertrand and Mullainathan (2004) and subsequent research has mostly interpreted the bias to be taste-based because, even when clear signals of productivity are added to the resumes sent to employers, discrimination is not eliminated, implying the evidence for statistical discrimination is weak. Our research uncovers evidence that workers may discriminate on the reciprocity dimension against their out-group employers by underperforming on the job. If employers develop beliefs about this bias and if reciprocal linkages are crucial to the positions they recruit into, then employers may exhibit belief-driven discrimination in their hiring decisions. In short, the employer-to-worker discrimination documented in the literature and mostly understood by economists to be taste-based may have a significant statistical component.

Our finding of altruism toward black employers is also of interest. We do not suggest it is omnipresent. Indeed, the prosociality may vanish in settings where employer-worker engagement is longer and involves physical interaction. Likewise, we recognize that unlike the current focus on the intensive margin of worker effort, understanding social preferences on the extensive margin may be equally important. After all, it is possible workers from the dominant group may systematically select out of (not even apply for) jobs posted by disadvantaged-group employers, thereby limiting the labor resources at the disposal of said employers. In short, if workers are given agency in who they work for, they may well avoid out-race employers. We aim to study the extensive margin angle to worker-to-employer discrimination in future research.

One concern about our results might be that worker subjects, knowing they are part of an experiment, maybe responding to experimenter demand effects or, more accurately, the Hawthorne effect. We contend that this concern is unlikely to be critical for the following reasons: 1) our treatment revelation mechanism is decidedly subtle, and taken in tandem with the between-subject design, it is almost impossible for subjects to pinpoint the real purpose of the study as being race-related; 2) even if there are demand effects they are likely to be mild – the recent paper de Quidt, Haushofer, and Roth (2018) finds that the range of (weak) demand effect for our 'a-b' task is expected not to exceed 11 percent of the treatment effects, which means the differential effort choices for black employer relative to the white, is likely to be between 85.33 and 95.27.[20] Taken together, these arguments imply that demand effects, if any, likely have a modest impact on our results.

We recognize that we do not offer a clear answer to the question, what explains either the pro-social behavior or the discrimination in reciprocity of white workers towards black employers? In our defense, we note that our study is a *first* attempt,

[20]Arguably, the demand effects in de Quidt et al. (2018) are not entirely driven by race. Relatedly, Mummolo and Peterson (2019) find, using a vignette study approach, that the experimenter demand effect in studies on racial discrimination is modest.

narrowly designed, to detect evidence. Of course, the larger question deserves full attention, and it will, in our future work.

References

Abel, M. (2019). *Do Workers Discriminate against Female Bosses?*

Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, *97*(4), 543–569.

Andreoni, J., Payne, A. A., Smith, J., & Karp, D. (2016). Diversity and donations: The effect of religious and ethnic diversity on charitable giving. *Journal of Economic Behavior and Organization.*

Arrow, K. J. (1973). The Theory of Discrimination. In *Discrimination in labor markets* (pp. 3–33).

Ayalew, S., Manian, S., & Sheth, K. (2018). *Discrimination from Below: Experimental Evidence on Female Leadership in Ethiopia.*

Becker, G. S. (1957). *The economics of discrimination.* University of Chicago Press.

Becker, G. S. (1974). A theory of Social Interactions. *Journal of Political Economy*, *82*(6), 1063–1093.

Benson, A., Board, S., & Meyer-ter Vehn, M. (2019). *Discrimination in Hiring: Evidence from Retail Sales.*

Bertrand, M., & Duflo, E. (2017). Field Experiments on Discrimination. In *Handbook of economic field experiments* (Vol. 1, pp. 309–393). North-Holland.

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*(4), 991–1013.

Cavaille, C. (2018). *Implementing Blocked Randomization in Online Survey Experiments.*

Chakraborty, P., & Serra, D. (2019). *Gender differences in top leadership roles: Does worker backlash matter?*

Charles, K. K., & Guryan, J. (2011). Studying discrimination: Fundamental challenges and recent progress. *Annu. Rev. Econ.*, *3*(1), 479–511.

Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual Behavior and Group Membership. *American Economic Review*, *97*(4), 1340–1352.

Czibor, E., Jimenez-Gomez, D., & List, J. A. (2019). *The dozen things experimental economists should do (more of).*

DellaVigna, S. (2018). Structural Behavioral Economics. In *Handbook of behavioral economics: Applications and foundations* (Vol. 1, pp. 613–723). North-Holland.

DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2016). Estimating social preferences and gift exchange at work. *NBER Working Paper Series*, *53*(9), 1689–1699.

DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*, *85*(2), 1029–1069.

de Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266–3302.

Dodlova, M., & Yudkevich, M. (2009). Gift exchange in the workplace. *Human Resource Management Review*, *19*(1), 23–38.

Doleac, J. L., & Stein, L. C. D. (2013). The visible hand: Race and online market outcomes. *Economic Journal*, *123*(572), 469–492.

Eckel, C. C., & Petrie, R. (2011). Face value. *American Economic Review*, *101*(4), 1497–1513.

Fehr, E., & Gächter, S. (1998). Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review*, *42*, 845–859.

Ferrary, M. (2003). The gift exchange in the social networks of Silicon Valley. *California Management Review*, *45*(4), 120–138.

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

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

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: resting for gift exchange in labor markets using field experiments. *Econometrica*, *74*(5), 1365–1384.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*.

Grossman, P. J., Eckel, C. C., Komai, M., & Zhan, W. (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior and Organization*, *161*, 197–215.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*.

Hitlin, P. (2016). *Research in the Crowdsourcing Age: A Case Study* (Tech. Rep.). Pew Research Center.

Kube, S., Maréchal, M. A., & Puppe, C. (2006). Putting reciprocity to work-positive versus negative responses in the field. *University of St. Gallen Economics Discussion Paper*.

Loury, G. C. (1998). Discrimination in the post civil rights era: Beyond market interactions. *Journal of Economic Perspectives*, *12*(2), 117–126.

Mummolo, J., & Peterson, E. (2019). Demand Effects in Survey Experiments: An Empirical Assessment. *American Political Science Review*, *113*(2), 517–529.

Neumark, D. (2018). Experimental Research on Labor. *Journal of Economic Literature*, *56*(3), 799–866.

Oh, S. (2019). *Does Identity Affect Labor Supply?*

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188.

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision making*, *5*(5), 411–419.

Phelps, E. S. (1972). The Statistical theory of Racism and Sexism. *American Economic Review*, *62*(4), 659–661.

Riach, P. A., & Rich, J. (2002). Field experiments of discrimination in the market place. *Economic Journal*, *112*(483), 480–518.

Rich, J. (2014). *What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000.*

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*.

Rotemberg, J. J. (2006). Altruism, Reciprocity and Cooperation in the Workplace. In S.-C. Kolm & J. M. Ythier (Eds.), *Handbook of the economics of giving, altruism and reciprocity* (pp. 1371–1407).

Simon, H. A. (1993). Altruism and Economics. *American Economic Review: Papers & Proceedings*, *83*(2), 156–161.

Sundstrom, W. A. (1994). The Color Line: Racial Norms and Discrimination in Urban Labor Markets, 1910-1950. *Journal of Economic History*, *54*(2), 382–396.