Causes and Consequences of Perceived Workplace Discrimination

Hannah Ruebeck*

October 2024

Abstract

Perceived discrimination is widespread but understudied. In two online experiments (N≈4,000), I randomly assign workers to be evaluated under procedures with varied potential to discriminate and the information workers have about past promotions. Learning that a manager knew race and gender and previously promoted mostly white men increases perceived discrimination among women and racial-minority men by 31pp, lowers women's retention by 5 percent, and increases reservation wages by 10 percent. Employers could improve equity and efficiency by reducing perceived discrimination. However, I find that implementing increasingly-common anti-bias hiring policies cannot alone eradicate perceptions of discrimination when minority groups remain under-represented.

JEL Codes: J7, D91, J53

*Hamilton College Economics Department. Email: hruebeck@hamilton.edu. This paper was previously circulated as "Perceived Discrimination at Work." I am immeasurably grateful for the support of my advisors Esther Duflo, Parag Pathak, and Frank Schilbach. I would also like to thank Abi Adams-Prassl, David Autor, Abhijit Banerjee, Katherine Coffman, Lisa Ho, Sendhil Mullainathan, Lucy Page, Ashesh Rambachan, Ray Reagans, Nina Roussille, Kartini Shastry, Broderick Turner, Lise Vesterlund, and the participants in MIT's labor and behavioral lunches for helpful discussion and feedback. This project was financially supported by the Social Policy and Research Initiative at the Abdul Latif Jameel Poverty Action Lab, the George and Obie Shultz Fund, Esther Duflo and Frank Schilbach's research funds, Hamilton College, and the Administration for Children and Families (ACF) of the United States Department of Health and Human Services (HHS). The latter was as part of a financial assistance award (Grant #: 90PD0314) totaling $25,000 (27 percent) funded by ACF/HHS and $66,913 (73 percent) funded by non-government sources. The contents are those of the author and do not necessarily represent the official views of, nor an endorsement, by ACF/HHS, or the U.S. Government. I was also supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. The two experiments were pre-registered as AEARCTR-0009592 and AEARCTR-0011806. IRB approval for both studies was obtained under MIT's Committee on the Use of Humans as Experimental Subjects as protocols 2201000547 and 2307001048. The *Online Supplement* can be accessed at https://sites.google.com/view/hannahruebeck/research

1 Introduction

In the last four decades, social scientists have documented widespread workplace discrimination (see e.g. Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017; Neumark, 2018; Small and Pager, 2020). Minority-group workers know that discrimination is pervasive, have experiences that could be due to discrimination, and often infer they have been discriminated against.[1] We know little, however, about either the causes or consequences of *perceived* discrimination because it is difficult to separate causally from discrimination itself.

Whether or not discrimination is present, perceived discrimination may exacerbate race and gender gaps if it affects worker behavior via distrust, retaliation, or anticipated further discrimination. Across many contexts, including in nationally-representative samples and my baseline survey, perceived discrimination is negatively correlated with job satisfaction, employer attachment, and psychological well-being (Appendix Tables A1 and A2 summarize this literature). Accounting for perceived discrimination can also explain well-documented gaps; for example, controlling for perceived discrimination closes Black-White gaps in labor supply (Goldsmith et al., 2004).

This paper describes two large, pre-registered experiments (N≈4000) that identify the causes and consequences of perceived discrimination in a simulated labor market.[2] The experimental design generates novel exogenous variation in perceptions of discrimination by randomly assigning workers to be evaluated under procedures with different potential to discriminate and telling rejected workers about how they were evaluated. I collect counterfactual evaluation decisions so that I can eliminate any differences in the sample of workers selected by each evaluation procedure.

The experiments address four research questions. The first experiment primarily tests whether perceived discrimination affects worker retention, effort, performance, and future labor supply. It also tests whether workers respond similarly to perceived manager and algorithmic discrimination, and whether improving minority-group representation can reduce perceived discrimination rates. The second experiment tests the effects of implementing common anti-bias hiring policies on *perceived* discrimination, holding minority-group representation fixed—a quicker and more straightforward potential solution for firms aiming to reduce perceived discrimination.

Both experiments take place in a labor market with multiple rounds of worker evaluation and work opportunities that I constructed on Prolific, an online platform where workers take surveys for pay. The short-term work relationship I study is most akin to the large "gig economy," which is

[1] 40 percent of women in the US report *ever* experiencing gender discrimination at work, and 25 percent of Black and Hispanic workers report experiencing discrimination at work *in the last year* (Pew Research, 2017; Gallup Inc, 2021). Among the workers screened into my first experiment, 20, 35, and 50 percent of Hispanic, Asian, and Black men and women report ever experiencing racial discrimination at work before the survey, and around 35 percent of white and racial-minority women report ever experiencing gender discrimination (Appendix Figure A3).

[2] This refers to restricted samples used in the main analysis; up to 5,000 workers are included in other analyses.

rife for (perceived) discrimination (Gallup Inc, 2018; Pew Research, 2021a,b). Workers in my simulated labor market proofread scientific texts, a task for which there exist gender and racial stereotypes that could engender perceived discrimination when a worker learns they were rejected.[3] The main experimental sample consists of white and racial-minority women and racial-minority men living in the US. To my knowledge, my study is unique in the lab-style experimental literature on discrimination in that it studies racial discrimination in addition to gender and is powered to detect heterogeneity along these lines; racial minorities most systematically experience discrimination in the workplace (Kline et al., 2021).

In the first experiment (N≈2,300), workers are recruited for an initial work task and one of three randomly-assigned procedures determines whether they are "promoted"—offered a harder, higher-paying job—or not, in which case they are offered an easier, lower-paying one. The procedures are (i) a *demographic-blind manager* who sees prior performance and education, (ii) a *non-blind manager* who also sees avatars, race, and gender, and (iii) an algorithm that predicts performance using prior performance and education. I implement the randomly-assigned procedure's decision for each worker. The analysis sample restricts to the 90 percent of workers who would not be promoted under *any* procedure, so the only differences between treatment arms are workers' perceptions of why they were not promoted.

After workers learn about how they were evaluated, I measure perceived discrimination and observe subsequent work behavior. I test whether perceived discrimination is quantitatively important by focusing on the effects of treatment on worker performance, retention, and future labor supply. I observe worker effort and performance in required pay-for-performance scientific proofreading tasks and the number of optional proofreading tasks they complete—which I call retention—and elicit incentivized reservation wages for future job opportunities, including ones involving cooperation with the managers who previously evaluated them.

I also introduce a new measure of perceived discrimination that circumvents concerns with other methods. Prior surveys prime workers by asking directly about discrimination, which could lead to over-reporting (e.g. Kessler et al., 1999; Goldsmith et al., 2004; Pew Research, 2017; Gallup Inc, 2021). Administrative data relies on filed complaints that may be under-reported. Instead, I ask workers what they think needed to be different about their profile to be promoted. I define per-

[3] I identify the race and gender stereotypes that participants hold about this task and each of its components (grammar, spelling, science knowledge) in a separate sample. There are systematically negative stereotypes against women and Black and Hispanic workers' performance in the scientific proofreading task and its components. There are positive stereotypes about Asian workers' performance. Perceived discrimination may be more common in domains with either positive or negative stereotypes (compared to neutral domains): when there are negative stereotypes, workers may be particularly likely to perceive statistical discrimination and when there are positive stereotypes, workers may be particularly likely to perceive taste-based discrimination. I do not separately identify perceived types of discrimination. I pool Black, Hispanic, and Asian workers as "racial minorities" due to power considerations; all are numerical minorities in the United States and large shares report having previously experienced negative racial discrimination in my sample (Appendix Figure A3).

ceived discrimination as whether a worker's open-ended response mentions demographics, which correlates strongly with more explicit questions asked at the end of the survey.

I first focus on the two manager arms to study the effects of perceived discrimination. Non-promoted workers see how they were evaluated before starting the job, generating differences in perceived discrimination. Workers see demographic and work-history information for their manager and the profiles of the three workers their manager promoted in the past, which communicates the information their manager had when making decisions. These previously-promoted workers are primarily white men due to their over-representation in earlier cohorts. Learning that one's manager knew workers' race and gender and previously promoted mostly white men—i.e. being in the non-blind manager arm—increases perceived discrimination by 31pp compared to the blind manager arm.[4] There are subsequently large effects on worker behavior.

Perceived discrimination affects retention and performance in ways that would exacerbate gender gaps and racial gaps among women. Being in the non-blind manager arm reduces the number of optional tasks that white and racial-minority women complete by 5 percent (p=0.01), whereas there is, if anything, a positive retention effect for racial-minority men. Racial-minority men also experience a *positive* performance effect in the required paragraphs, increasing earnings by 13 percent (p=0.04). Being in the non-blind manager arm seems to make racial-minority women try harder but do worse, while white women's performance is unaffected. These effects are not due to anticipated further discrimination, but reflect cognitive or emotional responses that change the perceived returns to, or cost of, effort.[5] The observed differences in effects by gender are consistent with documented differences in confidence in stereotyped domains and competitiveness—which I observe in my sample and which would lead to higher costs of effort for women and higher returns for men (e.g. Coffman, 2014; Bordalo et al., 2019; Niederle and Vesterlund, 2007). These and differences in beliefs about experiencing discrimination in the past and its prevalence in general explain some, but not all, of these heterogeneous effects.

Perceived discrimination also reduces future labor supply in ways that would exacerbate race and gender gaps relative to white men. Workers in the non-blind manager arm have 10 percent (p=0.03) higher reservation wages for a job opportunity with a chance for promotion that is determined by the same manager. I measure beliefs about the likelihood of future promotion and elicit reservation wages when there is no scope for additional discrimination, allowing me to disentangle these overall effects into: (i) lower expected wages due to anticipated discrimination and (ii) direct

[4]Workers are not told the demographic composition of these earlier cohorts. Testing whether my results would generalize to a setting in which workers knew these base rates is beyond the scope of the paper. Other work suggests that workers would still be inattentive to the demographics of the previously-evaluated cohorts when seeing that a white male manager previously promoted mostly white men (Benjamin, 2019; Hagmann et al., 2022).

[5]Workers could have anticipated discrimination in payment decisions, as they were potentially subjective. Empirically, however, workers anticipated that there were pre-determined correct answers and a cutoff rule to determine payment. At this point of the survey, they did not know there would be additional evaluation and job opportunities.

disutility from interacting with a biased manager. Taking point estimates at face value, 70 percent of the 10 percent increase in reservation wages is due to direct disutility rather than changes in expected wages. Workers in the non-blind manager group also have 23 percent (p=0.02) higher willingness to pay to be able to choose a different manager to work with in a collaborative task. The negative effects on future labor supply are similar for all workers. Thus, while there is heterogeneity in the cognitive and emotional responses to perceived discrimination when a worker is "stuck" in a job, everyone prefers to avoid future discrimination.

Workers in the non-blind manager arm are also 9 percent (p=0.10) less likely to share with their manager in a dictator game. Workers with more education, who believe discrimination is the most common, and who report ever experiencing discrimination are more retaliatory.

Next, I turn to the algorithm arm to understand whether workers' reactions to perceived algorithmic discrimination are as detrimental as their reactions to perceived manager discrimination. In addition to its policy relevance given the proliferation of algorithms in the labor market (Cowgill, 2020; Jarrahi et al., 2021), this comparison helps shed light on mechanisms: algorithmic bias is generated by statistical processes, not explicit prejudice, which may change how perceived discrimination affects behavior (Bigman et al., 2023). In the algorithm arm, workers are cross-randomized to learn that the (demographic-blind) algorithm previously promoted mostly white men, or to not learn the demographics of the previously-promoted workers at all. Learning that the algorithm previously promoted mostly white men increases perceived discrimination by 17pp.

Perceived algorithmic discrimination has different, but still negative, effects on worker behavior. Seeing that a demographic-blind algorithm previously promoted mostly white men reduces performance by 5-7 percent (p=0.09) but, if anything, positively affects retention. The effects are statistically significantly different from the effects of being in the non-blind manager arm. On the other hand, when workers expect additional discrimination, perceived algorithmic discrimination reduces future labor supply similarly to perceived manager discrimination.

Thus, the second part of the paper focuses on how firms might reduce the prevalence of perceived discrimination through their hiring or promotion policies. The experimental manipulations test how ongoing structural changes in the labor market might affect perceptions of discrimination. First, firms face growing pressure to diversify and de-bias their workforces, including by hiding worker demographics from decision-makers (Chang et al., 2019; Brecheisen, 2023; Fath, 2023; Gallup Inc, 2023). Second, as just described, algorithms are increasingly ubiquitous in managerial decision-making. All of these efforts may or may not lead to more minority-group representation, and even if they do, the effects on perceived discrimination are ambiguous.[6]

[6]Companies' increasing emphasis on diversity, equity, and inclusion may successfully change perceptions, or fail or backfire if they are inauthentic or ineffective. Flory et al. (2021) show that job ads emphasizing diversity increase minority-group application rates, regardless of whether the statement is evidence-backed, and Baker et al. (2023) show that firms opportunistically publicize diversity commitments and that these firms are *more* likely to incur EEOC

In addition to the design discussed so far, the first experiment also truthfully and randomly varies whether workers who see previously-promoted workers' demographics see a group that is all or two-thirds white men. This tests how changes in minority-group representation affect worker perceptions, holding decision-making procedures and inputs fixed. I compare these results to evidence from the second experiment (N≈2,700 new workers), which tests how anti-bias hiring policies affect perceived discrimination when the previously-hired workers are white men—i.e. the effects of changing decision-making *procedures* or *inputs*, rather than *outcomes,* when minority-group representation is minimal. The design largely follows that of the first experiment, but I randomly assign workers to be evaluated by a manager or an algorithm and cross-randomize whether the decision-maker knows their race and gender. Among workers evaluated by a demographic-blind decision-maker, I further cross-randomize whether they learn previous hires' demographics; they are all white men.[7] Perceived discrimination is the primary outcome, and workers do not do an effortful task—the sample is workers who are not "hired," rather than not "promoted."

Seeing one woman or racial minority man among the three previously-selected workers reduces perceived manager and algorithmic discrimination by 41-48 percent (p<0.005). The effect is largest when seeing a representative of one's own race and gender, but seeing a member of a different minority group also significantly reduces perceived discrimination rates. In contrast, changing hiring procedures to mitigate the direct effects of human biases—using algorithms or blinding managers to race and gender—does little to mitigate common *perceptions* of discrimination when past decisions "seem biased." When workers see that only white men were previously hired and know their manager knew their race and gender, 52 percent perceive discrimination.[8] Blinding managers to race and gender only reduces perceived discrimination by 11pp (20 percent, p=0.03) when white men are so over-represented. Regardless of whether decision-makers can use demographic information or not, using an algorithm *increases* perceived discrimination by 9-10pp (20-22 percent,

penalties and *less* likely to hire diverse candidates. Blinding decision-makers to demographics can improve representation of minority groups (e.g. Goldin and Rouse, 2000; Fath, 2023) or backfire (Behaghel et al., 2015). Increasing representation even moderately could reduce perceived discrimination (Brown et al., 2024; Kardosh et al., 2022) or not, if it is perceived as tokenism or there is backlash among majority-group workers (Anisman-Razin and Saguy, 2016; Leslie, 2019). Algorithms can embed human biases but are easier to regulate and audit, perhaps leading to less discrimination (e.g. Kleinberg et al., 2020) and thus less perceived discrimination, but they are also unavoidably opaque, undermining this potential. In any case, reducing actual bias may not reduce *perceived* discrimination, as reasoning about whether one has experienced discrimination is a complex problem. Individuals only receive few, noisy signals about why they themselves do not receive certain opportunities (Jones et al., 2016; Doering et al., 2023) but get many signals about the prevalence of discrimination at scale.

[7] This design separately identifies the effects of (1) using an algorithm and (2) blinding decision-makers when workers know that previously-promoted workers were white men. This is not possible in the first experiment: comparing the manager arms pools the effect of blinding managers to demographics and not sharing previously-promoted workers' demographics. Comparing the non-blind manager arm and algorithm sub-arm where workers learn previously-promoted workers' demographics pools the effect of using an algorithm and blinding the decision-maker.

[8] This is higher than in the first experiment because (i) all—not most—previously-hired workers are white men and (ii) women—who are more likely to perceive discrimination—make up a higher fraction of the second sample.

p<0.04). These results suggest that using algorithms or keeping demographic information from decision-makers (and telling workers about the procedural change) can, at best, only moderately reduce perceived discrimination if the outcomes of those procedures do not change.

In sum, this paper shows that the costs of discrimination may be larger than previously thought: the experience of feeling discriminated-against imposes additional material and psychological costs, adding to any direct harms of discrimination. Furthermore, employers could improve equity and efficiency by reducing perceived discrimination, but eliminating such perceptions will likely be unattainable as long as members of minority groups do not see themselves sufficiently represented in an organization. Implementing objectively race- and gender-neutral procedures cannot eliminate perceived discrimination when decisions seem intuitively biased. These concerns are likely to become even more central as opaque algorithms play an increasingly important role in rendering high-stakes decisions that affect access to opportunity.

This paper contributes to literatures in labor and behavioral economics and cross-disciplinary work on algorithmic bias.[9] First, racial and gender discrimination in the workplace is well-documented (e.g. Pager and Shepherd, 2008; Bartoš et al., 2016; Bertrand and Duflo, 2017; Blau and Kahn, 2017; Neumark, 2018; Bohren et al., 2019, 2022; Small and Pager, 2020; Kline et al., 2021). I show that perceived discrimination affects worker behavior and therefore exacerbates gaps caused by discrimination itself, including affecting future opportunities even if future evaluation is unbiased. Prior work therefore understates the overall effects of a discriminatory labor market. Glover et al. (2017) document similar self-fulfilling effects of discrimination—minorities exert less effort when they have biased managers—but their results are best explained by managers choosing to interact less with minority workers ("aversive racism"), rather than worker retaliation or avoidance after perceiving discrimination. Indeed, minority workers in their setting are *not* more likely to report that biased managers disliked them or they disliked biased managers. I document a type of self-fulfillment that could compound these effects and can occur even if managers are *unbiased*.

This paper's second contribution is to disentangle the effects of perceived discrimination into (i) changes in expected wages due to anticipated discrimination (ii) direct disutility from anticipated interactions with a biased manager. Psychological costs play a substantial role. Thus, I build on work showing the importance of integrating non-classical preferences into models of worker effort provision (Akerlof, 1982; Fehr et al., 2009; Card et al., 2012; Cohn et al., 2014; Breza et al., 2018; Dube et al., 2019; DellaVigna et al., 2022; Fehr and Charness, 2023).

In doing so, I also contribute to a theoretical and experimental literature on the effects of *anticipated* discrimination on human capital investment due to changes in incentives (Lundberg and

[9]The paper is also motivated by work in psychology, sociology, and economics showing that perceived workplace discrimination is highly predictive of many negative economic, social, and mental and physical health outcomes (see Appendix Tables A1 and A2 and Section 3.4.2 for details).

Startz, 1983; Coate and Loury, 1993; Fryer et al., 2005; de Haan et al., 2017; Dianat et al., 2020; Campos-Mercade and Mengel, 2024) and recent experiments showing that anticipated discrimination affects job search behaviors (Alston, 2019; Charness et al., 2020; Lepage et al., 2022; Ridley, 2022; Agüero et al., 2023; Aksoy et al., 2023; Avery et al., 2023; Angeli et al., 2024). These papers do not identify the mechanisms behind these effects and infer that workers' responses are entirely (and sometimes incorrectly) strategic, whereas I show that this avoidance may be due to psychological costs. Other work shows that workers value related concepts like dignity or lack of sexual harassment at work using vignette experiments and amenity valuation (Sockin, 2021; Dube et al., 2022; Folke and Rickne, 2022; Adams-Prassl et al., Forthcoming). I provide experimental evidence that perceived discrimination affects worker sorting into jobs and increases willingness to take a pay cut to avoid future discrimination.

Most closely related is contemporaneous work by Gagnon et al. (2024). In an experiment with Prolific workers in the UK, they show that explicitly telling workers that wage inequality is due to gender discrimination further reduces lower-wage workers' labor supply (analogous to the behavior I call retention) relative to the effects of wage inequality alone. In my setting, workers learn about hiring procedures and their results and possibly infer discrimination, providing new evidence on the causes of perceived discrimination; I use a novel measure to document these effects. With this, we learn that perceived discrimination reduces retention among those who naturally perceive discrimination from more realistic information provision.[10] I also show that perceived discrimination affects future labor supply and cooperation. They provide suggestive evidence that reduced morale drives their results; using various reservation wage and quantitative belief elicitations, I show that there are large psychological costs to perceived discrimination.[11]

Finally, an extensive literature in computer science, law, and economics shows that algorithms can either exacerbate or mitigate human bias (e.g. Kleinberg et al., 2018; Obermeyer et al., 2019; Kleinberg et al., 2020; Raghavan et al., 2020). There is also mixed evidence on how people perceive algorithmic versus human fairness, though there is little work on perceived race and gender bias specifically.[12] Two recent papers study revealed preferences for algorithmic versus manager decision-making and also find mixed results. Women are 35 percent more likely to apply to tech

[10] They also study inferred perceived discrimination by informing workers of the other worker's gender and wage and find that this makes workers (especially women) more likely to perceive discrimination as measured on a 7-point Likert scale in response to the question, "During the task, did you believe that gender discrimination was used to determine your payment per line." This treatment has no effect on workers' labor supply.

[11] There are two other notable differences. First, I consider the role of race in addition to gender and find substantial heterogeneity by race conditional on gender, as well as by gender. Second, my experiment tests these effects in a context in which there are gender and racial stereotypes about performance versus their neutral context.

[12] People find depictions of algorithmic or human decision-making more fair in different small-sample studies depending on the definition of fairness (Lee, 2018; Kaibel et al., 2019; Acikgoz et al., 2020; Newman et al., 2020; Noble et al., 2021; Zhang and Yencha, 2022). People are less morally outraged by descriptions of algorithmic discrimination than human discrimination due to inferred differences in underlying prejudice (Bigman et al., 2023).

jobs when they will be evaluated by an algorithm rather than managers because they anticipate less discrimination (Avery et al., 2023), but workers in a lab setting are about equally split when they can choose between manager and algorithmic evaluation (Dargnies et al., 2022). I experimentally compare perceptions of discrimination by managers and algorithms when workers can see that they previously promoted mostly white men. In this setting, workers are *more* likely to perceive algorithmic than manager discrimination, and both negatively affect worker outcomes.

The paper proceeds as follows: Section 2 describes the setting and research design. Section 3 provides evidence on the effects of perceived discrimination. Sections 4 and 5 describe the effects of of minority-group representation and anti-bias policies on perceived discrimination rates, respectively. Section 6 discusses robustness and Section 7 concludes.

2 Research design

The key challenge in studying the effects of perceived discrimination is that those who perceive discrimination are more likely to have experienced the direct effects of discrimination and to be from oft-discriminated groups, who are—observably and non-observably—different from others. Similarly, natural variation in firm policies changes the prevalence of discrimination alongside perceptions. This experiment overcomes these challenges by creating random variation in perceived discrimination while allowing me to account for any actual discrimination. Importantly, it does so while minimizing the impacts of any actual discrimination, being truthful with workers, and proxying a realistic work scenario in which workers are likely to perceive discrimination—circumventing new challenges introduced in an experimental context.

This section describes how I address these challenges. In summary, workers are randomly assigned to be evaluated by procedures that vary in their potential to discriminate. Workers who are not selected to do a difficult job learn about their evaluation and see workers that were previously selected under the same procedure, only sometimes learning these workers' demographics—they are primarily white men due to over-representation in past cohorts. Finally, the analysis sample restricts to workers who would not have been selected under any procedure, ensuring that differences across treatment arms come only from differences in perceptions of the procedures, and not the procedures' decisions themselves.

2.1 Setting and sample

I recruit workers on Prolific, an online platform commonly used in economics research where workers complete surveys for pay. Prolific is a suitable setting for my experiment for several reasons. First, manipulating perceived discrimination requires experimental variation and it is

costly and difficult to randomly manipulate worker-selection procedures in the field. My approach on Prolific, a functional labor market, allows me to vary perceived discrimination while holding fixed: (i) any confounding effects of actual discrimination, (ii) all information workers have about the jobs, selection procedures, and other workers, and (iii) the effect of not being promoted. I discuss differences and similarities with other employment contexts in Section 2.5.

I over-sample workers likely to experience and perceive racial and gender discrimination. 1,910 workers make up the main analysis sample in the first experiment, of which 55 percent are white women, 22 percent are racial-minority women, and 23 percent are racial-minority men. 170 white men otherwise qualify for the experimental sample but are used only as a comparison group to ensure that the treatments affect only perceived race and gender discrimination. 1,934 workers make up the main analysis sample in the second experiment, of which 72 percent are white women, 15 percent are racial-minority women, and 13 percent are racial-minority men. White women are more heavily over-represented due to constraints on the number of racial minority participants on Prolific who had not already participated in the promotion experiment. In both samples, the racial-minority men and women are approximately one-third each Asian, Black, and Hispanic.[13]

The experimental sample includes people who are representative of large swaths of the US population and 96 percent have non-Prolific work experience (Table 1). Seventy percent are currently employed outside Prolific. The sample is well-educated—50 percent have a 4-year college degree or more and 35 percent have some college experience but no degree or a two-year degree—and there is a large spread in income—about 50 percent have household income less than $60,000 per year, 25 percent earn between $60,000 and $100,000, and the rest earn above $100,000. The sample is on the young side of working age, with 70 percent between ages 18 and 44.

The experimental sample is more educated, but poorer and younger than a nationally representative sample. The share working outside Prolific is the same as the share of the US population in this age range in the labor force. Table 1 compares summary statistics from the experimental samples to their analogues from the nationally-representative 2021 ACS. The experimental sample is more likely to have a college degree, but also more likely to have household income between $20,000-$60,000 and less likely to earn above $160,000 than a representative sample.

[13] 3,240 [3,960] workers were recruited to the baseline survey for the first [second] experiment. 120 [360] are ineligible for the experimental sample because the list of possible managers to evaluate them was exhausted (note that randomization occurred within these groups of 120 [360]). Of these, 2,886 [3,510] are not promoted [hired] by their randomly-assigned evaluation procedure and thus offered the chance to participate in the experiment, which 80 [76] percent of them complete. Of those, 90 [94] percent would not have been promoted [hired] under any procedure regardless of their random assignment. The main analysis sample in the hiring experiment is further restricted to the random sample of workers who saw that only white men were previously promoted (76 percent).

2.2 The promotion experiment

The first experiment is referred to as the "promotion experiment" because workers are evaluated and assigned to one of two jobs, one of which represents "promotion" and the other "non-promotion." Non-promoted workers are the sample of interest and their performance in the non-promotion job is a key outcome. The second experiment is referred to as the "hiring experiment" because when workers are evaluated, they are either selected to do a job (the same as the "promotion job"), and thus "hired," or not. The non-hired workers are the sample of interest, and I do *not* observe their performance in an effortful job. It primarily follows the same structure as the promotion experiment with the deviations described in Section 2.3.

2.2.1 The jobs

Workers are offered scientific proofreading jobs. In the lower-paying, easier job ("non-promotion"), they proofread up to eighteen paragraphs from elementary-school-level articles and can earn up to $4.50 in bonuses. In the higher-paying, harder job ("promotion"), they proofread up to twelve paragraphs from articles in leading scientific journals and write a short summary of each, and can earn up to $9.00 in bonuses. Each paragraph has around 100 words and four inserted errors. Most workers use almost the entire allotted minute to proofread each paragraph. After the first six paragraphs, workers choose after each paragraph whether to continue proofreading.[14]

To proofread a paragraph, workers highlight words they think contain errors. They click on a word to highlight it, or again to un-highlight it if they change their mind. Their bonus is determined by the number of correctly-highlighted mistakes and incorrectly-highlighted non-mistakes when they submit their work: they earn the bonus for a paragraph if two times the number of correct highlights minus the number of incorrect highlights is at least the median of the same statistic for that paragraph in a pilot. Workers are told that they are paid for each paragraph in which they do "a good job" and that on average, they could expect to be paid for half of the paragraphs.

The scientific proofreading job was chosen for several reasons. First, of the effortful tasks that can be embedded in a survey experiment, it requires relatively high skill levels and effort and can be long without becoming tedious; it is also realistic work that might be completed online. Second, effort and performance are easily quantifiable, while performance is still potentially subjective. Since workers knew only that they would be paid per paragraph that they did a "good job" proofreading, workers could have anticipated further discrimination in payment decisions. Empirically, however, they anticipated the use of objective metrics.

[14] Workers who quit *before* finishing the first six paragraphs forgo their participation payment; 1.5 percent of potential participants who consented to taking the survey did so and this does not vary by treatment arm.

Third, scientific proofreading includes skills to which one might ascribe many different (positive or negative) stereotypes about gender or race, and the intention of choosing this task was to make those demographic characteristics feel potentially promotion-decision relevant. Given the combination of skills required in this task and various demographic groups, however, what stereotypes prevail is not clear *ex ante* and may matter for the generalizability of my results. I collected incentivized beliefs about these stereotypes in a separate sample (detailed in Section 2.5).

What did workers know about the jobs? At the beginning of the experimental survey, workers are told that they are working on a team that proofreads scientific articles and a [manager/algorithm] assigned workers to one of two jobs. They learn what makes one job harder and the other easier—the complexity—and that the harder job pays two to three times as much per paragraph (see the *Online Supplement* for the specific implementation).[15] They then learn that they were assigned to the easier job.

2.2.2 Design overview and background

Figure 1 illustrates the overall experimental design. Workers are randomly assigned to be evaluated by one of three promotion procedures with varied potential to discriminate. These procedures—managers who do or do not know workers' race and gender and a demographic-blind algorithm—know workers' prior performance and education and decide who to promote. Before recruiting the experimental sample, the same managers and algorithm evaluate groups of workers in which white men were over-represented (a timeline is in Figure 2). The key to inducing variation in perceptions of discrimination in the experimental sample is that workers learn about the procedure under which they were not promoted and who was previously promoted under that procedure.

The historical sample. The experiment aims to proxy a setting in which a worker from an under-represented group is not promoted and knows that mostly white men were previously promoted. To generate these dynamics online, 1,800 predominantly (85 percent) white male "historical workers" are initially evaluated by the *same managers* who later evaluate the experimental sample, and by the same algorithm. The experimental workers see who their manager or the algorithm previously promoted when they learn about their own evaluation.

The managers. The managers, all white men,[16] are recruited from Prolific and tasked with choosing three of twenty-four historical-sample workers to promote. They know their decisions will be implemented and the performance of the workers they promote will determine their bonus

[15] At the end of the survey, 68 percent of workers report preferring the harder job, 6 percent were ambivalent, and 20 percent were unsure but didn't think they would prefer the harder job. This does not differ by treatment group.

[16] This serves two purposes: to best represent cases in which a women or racial minority man might be the most likely to feel discriminated against and to minimize noise in the experiment caused by variation in manager characteristics (only 54 [20] participated in the promotion [hiring] experiment).

(most of their total payment). Half of the managers are randomly assigned to be "demographic-blind" (they see only baseline performance and education levels) while the other half are "non-blind" and also see workers' race, gender, and an avatar that workers created in the baseline survey. Baseline performance is shown as 1-5 stars indicating the quintile of workers' average quiz scores and education as no college degree, college degree, or more than a college degree.

The same managers return to evaluate workers in the experimental sample and maintain the same "demographic-blind" or "non-blind" status as when they evaluated the historical sample. They evaluate workers with similar quiz scores as those they evaluated in the historical sample. Each manager evaluates three sets of forty workers independently, promoting three from each set. Their decisions are implemented for one randomly-chosen set, and the performance of workers they promote in that set determines their bonus payment.

The algorithm. The algorithm uses a random forest regression-based model with workers' quintile of average quiz scores and education as inputs to predict performance in the harder job. It promotes the three workers with the highest predicted performance in each group of twenty-four (in the historical sample) or forty (in the experimental sample). It was trained on a sample of 500 workers recruited from MTurk using Cloud Research's pre-approved participant pool in which women and racial-minority men are also over-represented.

The *Online Supplement* further describes the managers and algorithm and their decision-making.

2.2.3 Experimental design

This section describes the promotion experiment design using the vocabulary from the previous section (*algorithm*, *demographic-blind* or *non-blind manager,* and *previously-promoted workers*, which refers to promoted workers in the historical sample). Figure 3 illustrates the treatment arms.

Baseline survey. Workers are recruited on Prolific for an initial survey, during which they take three timed quizzes that are predictive of their ability in the proofreading jobs—a spelling quiz of scientific words, general grammar questions, and general-knowledge science questions—and report how well they thought they did on each, the basis of an incentivized measure of confidence in each domain. Workers know their performance on the quizzes could affect their future work opportunities and are paid a small bonus for each correct answer. Workers then answer demographic questions, build an avatar that "looks like them,"[17] report a detailed work history, including past experiences of discrimination at work, and report incentivized beliefs about the prevalence of discrimination against different groups.

Randomization and worker evaluation. Workers are randomly assigned to groups of 120

[17]Workers made their avatars following the procedure shown in Appendix Figure A13. They did not know what the avatars will be used for or that they would be seen by other participants. They largely choose avatars that align with their self-reported race and gender in expected ways (Appendix Figure A14).

with average baseline quiz scores within one quintile of each other. Each group is evaluated by two randomly-assigned managers—one demographic-blind, one not—and the algorithm. To ensure that low-scoring workers might perceive discrimination, groups were paired with managers who evaluated workers in the same two quiz-score quintiles in the historical sample.[18]

Each worker has an equal probability of being randomly assigned to have the decision of the demographic-blind manager, non-blind manager, or algorithm implemented. Among workers assigned to be evaluated by the algorithm, half are randomized to see previously-promoted workers' avatars, race, and gender in their profiles (like the non-blind manager arm) whereas the other half do not (like the demographic-blind manager arm). Workers' baseline characteristics are balanced across arms (Appendix Table A3).

Workers that are not promoted under their randomly-assigned promotion procedure are offered the easier job; 80 percent take it up and become the experimental sample (N=2,397). They are all offered the same Prolific task and only learn about their evaluation *after* starting the survey. Workers are invited to participate in the experimental survey at least three weeks after the baseline survey. During this time, managers returned to evaluate the experimental sample; the delay also ensured that workers did not connect the questions about discrimination in the baseline survey to the study topic.[19] Attrition after starting the survey was uncommon, and similar across all treatment arms (Section 6). Promoted workers are offered the harder job.

Information about non-promotion and evaluation procedures. Upon starting the survey, workers learn about the two jobs and that they were not promoted. Framed as context for the next question*, "what do you think needed to be different about your profile to be assigned the harder job?,"* they then learn how they were evaluated.

In both manager arms, workers learn that their manager was recruited from Prolific, reviewed example paragraphs, evaluated worker profiles, and was paid based on the performance of the workers they promoted (and so should have been trying to promote the best workers). They see their manager's profile (avatar, age, education, gender, race, and tenure on Prolific), so differences in perceived discrimination between the two manager arms control for having a white male manager. The differences in what workers know about their evaluation comes next: they learn that their manager previously evaluated workers with similar quiz scores as them, then see their own and the previously-promoted workers' profiles. The profiles communicate what the manager knew when making their decision, i.e. whether they saw workers' race, gender, and avatar (Figure 4).

Thus, "treatment" in the manager sample is learning that one's manager knew race and gender

[18]This stratification ensures that workers see previously-promoted workers with similar scores to themselves and can infer they had a chance at the job. Absent this, low-scoring workers would mainly see top-scoring workers being promoted and would be unlikely to attribute decisions to discrimination. This design component was successful in ensuring that perceived discrimination was not highly correlated with performance (Appendix Figure A15).

[19]Less than 1 percent of the control group thought the topic was discrimination at the start of the experimental survey.

and seeing the avatars, race, and gender of previously-promoted workers (mostly white men), relative to the "control group" whose manager did not know race and gender and who does not know that the previously-promoted workers are mostly white men.

In the algorithm arm, workers learn that an algorithm used data from previous iterations of the survey to predict who would do the best at the harder job. They are explicitly told that the algorithm only uses information on average quiz scores and education to predict performance. Next, workers see previously-promoted workers' profiles. Half are randomly assigned to see profiles with avatars, stars, education, gender, and race; the other half see only stars and education (Appendix Figure A1). Thus, "treatment" in the algorithm arm is seeing the avatars, race, and gender of previously-promoted workers (mostly white men), relative to the "control group" who does not know that the previously-promoted workers are mostly white men.

Finally, when workers see previously-promoted workers' avatars, race, and gender, 40 percent of workers see that three white men were previously promoted. The rest see two white men and someone from another demographic group. Conditional on the type of evaluation used, this is random due to the random pairing of workers with managers or groups of historical-sample workers jointly evaluated by the algorithm.[20]

Eliciting perceived discrimination. After learning about their evaluation procedure and who was previously promoted, workers are asked, *"what do you think would have needed to be different about your profile for you to be assigned to the harder job? For example, would it have helped if you scored higher on the quizzes, or had more education?"* and answer in a text box. The main measure of perceived discrimination is an indicator for whether workers' responses suggest that they think their demographics (age, race, gender, etc.) played a role.[21]

To my knowledge, no other study has attempted to identify individuals who felt that they had been discriminated against without asking directly.[22] This measure aims to capture that feeling without priming workers to think about discrimination or their own race and gender. Other, more explicit questions—asked at the end of the survey and described below—correlate highly with the main measure (Appendix Figure A2) and results are similar using any measure of perceived discrimination. Detailed variable definitions are in the *Online Supplement*.

Observed work outcomes. Workers are required to proofread six paragraphs (spending a

[20] Workers were randomly paired to managers but other manager characteristics could correlate with whether they promoted only white men in the historical sample. Empirically, there are no systematic correlations.

[21] This was pre-registered as the main measure of perceived discrimination. The free response was coded as mentioning demographics independently by two MIT undergraduates and a PhD psychologist who professionally codes qualitative data. The externally-generated variables used in the analysis are highly correlated with variables generated by the author. The full coding scheme is in the *Online Supplement.*

[22] Surveys (e.g. Pew Research, 2017; Gallup Inc, 2021) ask whether someone has experienced discrimination or been treated differently because of their gender or race, "ever" or in the past year; see Appendix Tables A1 and A2 for other Likert-style scales used in the sociology and psychology literatures.

maximum of six minutes) to earn their participation payment; after the sixth and each subsequent paragraph they are explicitly asked if they want to proofread another paragraph or skip to the end of the survey. The primary outcome is how many paragraphs workers choose to proofread, including indicators for whether they did more than six or all eighteen.

In each paragraph, workers highlight (or unhighlight) identified errors by clicking on a word. For each paragraph, I observe the time spent, the number of clicks, and how many final highlights correctly identify mistakes or incorrectly identify non-mistakes. A summary measure of performance is their earned bonus. After the first, third, or fifth paragraph—timing randomly assigned—workers also complete a psychological scale eliciting emotional states.

Revealed-preference survey outcomes. After workers finish the proofreading task, I elicit reservation wages via a multiple price list (MPL): workers indicate whether they would like to participate in a future round of the survey at various wage schedules for the harder and easier job; the wage in the harder job is always twice the wage in the easier job. "Participation in a future round" involves being evaluated again and either being promoted or not promoted. One wage schedule is randomly chosen and workers' choices are implemented for a random subset of workers. Workers answer these questions under two conditions: if they would be evaluated by the same procedure as in the experiment and if their assignment would depend only on whether they had the highest baseline quiz scores. After each MPL, they report beliefs about the probability that they will be promoted if they participate with the given evaluation procedure.

In the manager arms only, I also elicit incentivized reservation wages for a collaborative task with the same manager and workers' willingness to pay to be able to choose a different manager to work with. Workers in the manager arms also play a dictator game. They know they may be randomly chosen to receive a $20 thank-you bonus and choose how much they would want to share with their manager if they are selected, in which case their choice will be implemented.

Self-reported outcomes. Workers report their interest in future work, job satisfaction, and whether they would prefer the harder or easier job in the future, and complete self-efficacy scales for the proofreading jobs and related skills.

Secondary measures of perceived discrimination. Finally, they answer a series of questions that validate the main measure of perceived discrimination. First, they are asked if they have a complaint about the promotion procedure and if so, they can describe their complaint in a open-ended response, which is coded for complaints about discrimination using a similar procedure to the main measure. Second, they answer multiple-choice questions about whether they would have been promoted if they were a different race, or a different gender (masked with similar questions about education and quiz scores); if they say "Yes, I think so" or "Yes, definitely" they are coded as perceiving discrimination. Finally, they answer incentivized questions about whether they thought various groups of workers were over- or under-represented (or neither) among promoted workers,

a measure of perceived discrimination in *general* rather than against themselves specifically.

2.3 The hiring experiment

The hiring experiment tests how hiring policies affect whether workers perceive discrimination. Specifically, it identifies the effects of using an algorithm versus a manager and the effect of blinding decision-makers to race and gender when workers see that primarily white men were previously hired. It takes place in the same constructed online labor market with newly-recruited but similar workers. The not-hired workers take an experimental survey without any proofreading job, and the key outcome is perceived discrimination.

3,960 new workers take a shortened baseline survey and then are evaluated by various procedures. There are six treatment arms; four replicate the arms of the promotion experiment (Figure 5). Workers are randomly assigned to be evaluated by a manager or an algorithm and cross-randomized to arms where the decision-maker knows race and gender or not. Workers evaluated by a demographic-blind decision-maker are cross-randomized again to either see the avatars, race, and gender of previous hires or not. Again, workers are evaluated by all four procedures, but the decision of their randomly-assigned procedure is implemented. Workers are balanced across treatment arms (Appendix Table A4).

The *same* managers evaluate these workers as in the promotion experiment. They maintain their earlier random assignment as demographic-blind or non-blind, and now choose one worker from groups of forty to do the harder proofreading job. All other workers are offered the experimental survey without any proofreading job.

The demographic-blind algorithm is the same as in the promotion experiment. The non-blind algorithm adds interactions of each baseline performance and education predictor variable with the eight race $\times$ gender groups. I aim to proxy a realistic hiring algorithm which would use sufficient predictors to be able to approximate race and gender. The non-blind algorithm does not observably discriminate—it meets the same calibration criteria as the demographic-blind algorithm and in fact, would hire more racial minorities than the demographic-blind algorithm.

When non-hired workers return for the experimental survey, they learn about the proofreading job, are told they were not chosen for it and see the analogous information about the managers and algorithms as in the promotion experiment. They then answer the question about what needed to be different about their profile in order to be hired. They do *not* subsequently do a proofreading job, but continue directly to reservation wage elicitations. Finally, they answer questions measuring memory and comprehension of the evaluation procedures and the questions that underly the secondary measures of perceived discrimination.

2.4 Experimental fidelity: Attention and comprehension

The mechanism ensuring attention to the information about their evaluation was the open-ended question, *"what do you think needed to be different about your profile to be assigned to the harder task?"* Workers had to answer this question; fewer than 5 percent gave an unexpected reason.

In the hiring experiment, workers also answered comprehension questions about the procedures at the end of the survey. Around 80 percent of workers in the manager arms knew the managers' payment structure and around 70 percent of workers in the algorithm arms knew that the algorithm predicts performance. 88-98 percent of workers knew that the decision-makers did use education and average quiz-score quintiles represented by stars and did not use age, work history, or how long they've worked on Prolific; this did not vary by treatment group.

In both experiments, workers answered an "attention check" during the final survey section with unincentivized outcomes. 75 and 80 percent of workers answered the question correctly in the promotion and hiring experiments, respectively. There is no differential attention across treatment arms and the results are robust to dropping the workers who failed the attention check (Section 6).

2.5 Generalizability

2.5.1 The Prolific labor market

As described earlier, my constructed labor market on Prolific provides the control needed to identify the causes and consequences of perceived discrimination without the effects of discrimination itself biasing the results. Of course, this constructed labor market differs in many ways from traditional workplaces; it also, however, shares many features with other contexts of interest. Most importantly, workers are paid for their effort and performance in a cognitively demanding task. While it is a short-term working relationship—the average (median) worker spent 38 (33) minutes on the experimental survey—it is much like those characterized by the growing "gig economy." About one in six Americans has participated in app-facilitated gig jobs that tend towards a similar duration. Racial minorities are over-represented in this kind of work and perceived discrimination may be common; 41 percent of racial minorities say they have felt unsafe while completing these jobs (compared to 28 percent of white workers) and 22 percent of women (15 percent of men) report facing unwanted sexual advances on the job (Pew Research, 2021a,b).

That said, the nature of the manager-worker relationship in my experiment has unique features even among short-term arrangements, in particular the limited "interaction" between managers and workers. Managers evaluate workers without any direct communication and workers learn of managers' decisions with limited information about their manager. This is more akin to perceived discrimination in resume screening, rather than an interview process or internal promotion

decision. In my baseline survey and more representative samples, perceived discrimination in job search is more common than in promotion decisions, but perceived discrimination in day-to-day activities ("micro-aggressions") is the most common (Appendix Figure A3, Kessler et al., 1999; Goldsmith et al., 2004). It is ambiguous how my results might change in a setting with repeated interactions, so this is an interesting avenue for future work.[23]

A final potential concern is that workers knew they were participating in a research study, and trust in an employer may be different from trust in a researcher-employer. Again, the role of trust in people and institutions is an important topic for future research in this area.

2.5.2 Stereotypes about performance in the scientific proofreading task

The scientific proofreading task was chosen, in part, because it is composed of tasks associated with common gender and racial stereotypes and I anticipated that workers would be more likely to perceive race or gender discrimination in a domain in which demographics feel decision-relevant.

Which stereotypes workers hold about performance in the scientific proofreading task may matter for the generalizability of my results. I elicited incentivized beliefs about performance gaps for different demographic groups in a separate sample of 517 Prolific participants, including second-order beliefs among women and racial-minority men (like the experimental sample) about what beliefs they think white men (like the managers) hold.[24] I elicit beliefs about performance gaps in the basic grammar quiz, science knowledge quiz, spelling quiz of scientific words, and scientific proofreading task and second-order beliefs for the scientific proofreading quiz. The survey sample, design, limitations, and detailed results are described in the *Online Supplement*.

Gender stereotypes. Gender stereotypes do differ by domain, but participants are most likely to anticipate that men out-perform women in all four tasks. The stereotype is strongest for science knowledge and weakest for grammar, with spelling scientific words and scientific proofreading falling (rationally) in the middle. Women and racial-minority men anticipate that white men strongly believe that men would out-perform women in the scientific proofreading task, more so than they themselves think; women's first- and second-order beliefs are only very weakly correlated. 82 percent of workers either themselves believe that men out-perform women in the scientific proofreading task or expect that white men think so (Appendix Table A5). Thus, both the first-and second-order gender stereotype beliefs that could affect worker behavior in my experiment suggest that my results are most relevant to male-typed domains.

[23] Workers may perceive more discrimination if rejection is more personal or less if they have more context. Effects on behavior may be larger in an interactive setting if they are due to retaliation or distrust, or smaller if there are higher stakes or more social or organizational supports.

[24] This survey was run one (two) year(s) after the hiring (promotion) experiment and was not pre-registered; IRB approval was obtained under Hamilton College's IRB with ID number F24-002.

Racial stereotypes. Beliefs about performance gaps are remarkably consistent across all four domains for Black, Hispanic, and Asian workers compared to White, Non-Hispanic workers. There is a strongly-held belief that Asian workers out-perform White, Non-Hispanic workers: 95 percent of participants themselves believe or expect white men to believe that Asian workers perform better than White, Non-Hispanic workers in the scientific proofreading task. Participants believe that there are negative stereotypes against Black and Hispanic workers, but these are not as strongly held as the negative gender stereotype or the positive Asian stereotype: about two-thirds of participants either believe themselves or expect white men to believe that Black or Hispanic workers under-perform White, Non-Hispanic workers (Appendix Table A5).

Either positive or negative stereotypes are likely to generate more perceived discrimination upon rejection than there would be in a neutral domain. My analysis focuses on the effects of perceived discrimination, regardless of its cause, but future work should try to understand more precisely whether stereotypes affect perceived discrimination and its effects. In the analysis below, I pool Asian, Black, and Hispanic workers as racial minorities, as analyzing the effects on these groups separately is underpowered. All are numerical minorities in the United States and workers in all three groups report experiencing (negative) racial discrimination at work in the past (Appendix Figure A3) and at similar levels in the experiment (Appendix Figure A4).

3 The effects of perceived discrimination

3.1 Estimation

Perceived manager discrimination. To first understand the effects of perceived discrimination in a status quo setting, I focus on intent-to-treat (ITT) specifications that compare outcomes between workers assigned to the demographic-blind manager arm and the non-blind manager arm in the promotion experiment. I estimate the following:

$$Y_i = \alpha + \beta T_i + X'\delta + H'\lambda + \gamma_g + \varepsilon_{ig} \tag{1}$$

among workers evaluated by a manager, where T_i is one if worker i is in the non-blind manager arm and zero otherwise; β is the effect of learning that one's manager saw avatars, race, and gender and previously promoted mostly white men. γ_g is fixed effects for workers' quiz-score quintile group, X is a vector of baseline worker characteristics and H is a vector describing the education and quiz scores of the previously-promoted workers seen by worker i.[25] The *Online Supplement* provides

[25] In the main specification, X includes the pre-registered demographic controls: baseline quiz scores, education, income, age, employment status, gender, and race. H includes indicators for whether the worker saw one, two, or three previously-promoted workers with the maximum number of stars possible in their quiz-score group (the rest

details on the measurement of all control variables. The choice of controls does not affect the main results (Section 6). Standard errors are robust to heteroskedasticity.

Perceived algorithmic discrimination. Next, I test whether perceived algorithmic discrimination has the same effects. To do this, I use analogous ITT specifications in the algorithm arm of the promotion experiment, comparing outcomes between workers who were cross-randomized to see previously-promoted workers' race, gender, and avatar and those who did not. Now, T_i in Equation 1 indicates seeing the demographics of previously-promoted workers; β is the effect of learning that the demographic-blind algorithm previously promoted mostly white men.

Sample restriction. One threat to the causal interpretation of β as the effect of an intervention that changes only *perceptions* of discrimination is that different promotion procedures could have promoted different types of workers. Since the managers' and algorithm's real decisions were implemented, this could generate differences between the treatment arms among workers who are not promoted. While the sample size is small, the different procedures do appear to select systematically different workers: for example, the non-blind manager selects fewer white men and women than the blind manager or algorithm, and the managers put more weight on having more than a college degree than the algorithm. To ensure that the only differences between arms were workers' perceptions, the experimental design builds in the counterfactual evaluation of every worker by every procedure. In the main analysis, the sample restricts to workers who would not have been selected regardless of their random assignment—the workers not promoted under any procedure. In the promotion experiment, this is 2,080 workers of 2,317. That said, the main results are similar when I use the restricted or full sample (Section 6).

An instrumental variables (IV) specification. Results that instrument for the main measure of perceived discrimination with random assignment are in the *Online Supplement*. These effectively divide the ITT estimates of the effects on behavior by the "first stage" effect on perceived discrimination. The IV estimates can be interpreted as the effect of perceived discrimination, but to do so, random assignment must only affect outcomes via its effects on *the instrumented measure of perceived discrimination*. This assumption may be violated—for example, only a certain degree of perceived discrimination may have effects, and this may be more or less frequent than the measures of perceived discrimination I collect. Of those I collect, the main measure has the largest first stage, implying the lowest IV estimates. There may be workers who perceive discrimination and are affected that are not picked up by this measure, in which case the IV estimate may still be an over-estimate. The ITT estimates are thus the focus in the paper; they represent a lower bound on the average effects of perceived discrimination caused by my treatments.

had one fewer star) and ten exhaustive and mutually-exclusive indicators for how the experimental-sample worker's education compared to the education of the previously-promoted workers they saw. The controls in H were not pre-registered but do differ by treatment group; the results are the same excluding these controls (Section 6).

3.2 Manipulation check: Treatment affects perceptions

Perceived manager discrimination. Workers' perceptions of discrimination strongly depends on whether they learn that their manager knew demographic information and previously promoted mostly white men. Figure 6 plots rates of perceived discrimination in each manager arm estimated using Equation 1 for the full sample and separately by racial minority status × gender.[26]

Perceived discrimination is uncommon in the demographic-blind manager arm. Only 3 percent of workers mention demographics when asked about what needed to be different about their profile in order to be assigned the higher-status job; these workers' open-ended responses indicate that they think the manager had more information than was shown in the profiles.

Overall, being in the non-blind manager arm increases the share of workers perceiving discrimination by 31pp (se=2pp).[27] Being in the non-blind manager arm increases perceived discrimination by 35pp (se=3pp) for white women, 41pp (se=5pp) for racial-minority women, and 19pp (se=4pp) for racial-minority men. The effect for white and racial-minority women is not statistically significantly different (p=0.30), but the effect for racial-minority men is significantly smaller than the effect for racial-minority women (p<0.001). Appendix Figure A4 plots rates in each treatment arm for each of eight race × gender cells; women generally perceive discrimination at similar rates regardless of race, as do racial-minority men.[28]

Results are similar using the secondary perceived discrimination measures (Appendix Table A6). The magnitude of treatment effects vary, but there is always a large, significant effect of being in the non-blind manager group and patterns across demographic groups are the same.

White men are not significantly more likely to perceive discrimination in the non-blind manager arm than the blind manager arm. The few who do seem to have either perceived discrimination on the basis of age (via gray hair) or been confused. This confirms that, as designed, I am studying

[26] Heterogeneity by race and gender was pre-registered, along with heterogeneity by past experiences of discrimination, beliefs about the prevalence of discrimination, and confidence. Gender and race most systematically predict perceived discrimination, but multivariate regressions with all pre-registered characteristics and other demographics are in Appendix Figure A16.

[27] Almost all workers who do not perceive discrimination cite needing more education or higher quiz scores to be promoted. Being in the non-blind manager arm reduces the share who cite the *primary* reason being their education or quiz scores by 11pp and 6pp (26 and 12 percent, se=2pp and 3pp), respectively (Appendix Table A17). The main measure of perceived discrimination is an indicator for mentioning demographics *at all*; being in the non-blind manager arm increases the share of workers who mention demographics as the *primary* reason they think they were not promoted from 1 to 19 percent (18pp, se=2pp, Appendix Table A6).

[28] Why are there such stark gender differences? One potential explanation for the differences in perceived discrimination rates is that women are seeing primarily (white) men being hired in an environment where they believe that men are expected to outperform them, whereas racial minority men see white men being hired in an environment where they either believe they are expect to outperform white men (Asian men) or there are weaker negative stereotypes than for women (Black and Hispanic men); see Section 2.5.2. However, Asian, Black, and Hispanic men perceive discrimination at similar rates. Aksoy et al. (2023) similarly find that women are more likely to anticipate discrimination on the basis of sexual orientation than men.

the effects of perceived racial or gender discrimination among racial minorities and white women. Therefore, in the subsequent analysis, I drop the 94 white men from the manager sample when estimating Equation 1. This was not pre-registered, but does not matter for results (Section 6).

Perceived algorithmic discrimination. Worker perceptions also depend on whether they learn that the (demographic-blind) algorithm previously promoted mostly white men. No workers perceive discrimination when they do not see the avatars, race, and gender of previously-promoted workers and are evaluated by a demographic-blind algorithm.[29] Learning that the algorithm previously promoted mostly white men increases perceived discrimination rates by 17pp (se=3pp) (Figure 7). The effect is 18pp for white women (se=4pp), 26pp (se=7pp) for racial-minority women, and 9pp (se=4.5pp) for racial-minority men.[30] Again, the effects do not differ by race among women (p=0.30) but among racial minorities, they differ by gender (p=0.02). Confirming the treatment manipulation, white men do not perceive discrimination and are dropped from the sample (N=76).

These treatment manipulations were *not* designed to provide evidence on the causes or prevalence of perceived discrimination, but to move perceived discrimination as much as possible to identify its effects on worker behavior. In the manager arms, the treatment pools learning that one's manager knew demographic information with learning that mostly white men were previously promoted. Comparing rates of perceived discrimination across the non-blind manager arm and algorithm arm when workers learned previously-promoted workers' demographic information pools the effects of using an algorithm, blinding the decision-maker, and the fact that the algorithm is less less likely to promote three (versus two) white men in the historical sample. Sections 4 and 5 decompose the causal effects of each of these components on worker perceptions in turn. Relatedly, I do not provide any information about the underlying demographics of the previously-evaluated cohorts in which white men were very over-represented. Testing the effects of revealing this information is beyond the scope of this paper but deserves attention in future work.[31]

[29] This was not obvious *ex ante*, given high-profile articles like 'Amazon scraps secret AI recruiting tool that showed bias against women', Reuters. Participants could have inferred bias just from the word "algorithm."

[30] Results are similar using the secondary measures of perceived discrimination (Appendix Table A18).

[31] Most workers in the US presumably have information about the demographic composition of their occupation and industry, so the high prevalence of perceived discrimination in the population (Pew Research, 2017; Gallup Inc, 2021) suggests many workers may still perceive discrimination if they were provided with this information. A large literature on base-rate neglect and the representativeness heuristic (Benjamin, 2019) further suggests that some workers in my experiment would likely still perceive discrimination in the non-blind manager arm with this information. In the context of hiring, Hagmann et al. (2022) show that people largely ignore the demographic composition of a worker pool when making inferences based on a selected few. This is discussed further in Section 5.

3.3 Effects of perceived manager discrimination on effort and performance

Perceived discrimination does not affect retention and performance overall, but has negative effects for women and positive effects for racial-minority men.

3.3.1 Effects of being in the non-blind manager arm

Retention. Workers must proofread six paragraphs to be paid their participation wage, and can then choose to proofread up to eighteen. On average, learning that one's manager knew race and gender and previously promoted mostly white men—being in the non-blind manager arm—has no statistically significant effect on retention, though the point estimates are negative (Table 2).

Intensive effort and performance. I identify causal effects of treatment on effort and performance in the first six required paragraphs, before treatment can cause differential selection in or out of the sample. Being in the non-blind manager arm does not affect measures of how hard workers choose to work: time spent per paragraph (capped at 1 minute) or the number of times they click on the page (a proxy for whether they check their answers). On average, it *increases* the number of correct highlights in the first six paragraphs (Table 3).

3.3.2 Heterogeneity by gender and race

Organizing framework. These effects of perceived discrimination on retention and performance are expected to operate through cognitive or emotional responses that change the perceived returns to, or cost of, effort, and there are reasons to think that these responses vary systematically by race and gender. Workers did not anticipate manager discretion in determining their payment for each paragraph, so perceived discrimination is not operating through the anticipated *monetary* returns to effort.[32] Thus, changes in the returns to or costs of effort must be psychological. Captured attention or stress responses might make effort more cognitively costly (Duquennois, 2022; Schmader et al., 2008), or lowered confidence or motivation might reduce perceived returns to effort. A physiological "challenge" response might make effort less cognitively costly (Vick et al., 2008), or a desire to prove the manager wrong might increase the perceived returns.

Perceived discrimination may affect these returns to and costs of effort (and therefore retention, intensive effort, and performance) in a way that exacerbates existing gender and racial gaps. For example, gender differences in competitiveness and confidence are known to cause women and

[32]This was intentionally left ambiguous, but workers did not think this was the case. Towards the end of the survey, workers were asked the open-ended question, *"how do you think it is determined whether you did a "good job" [and thus are paid for] proofreading these paragraphs?"* Fewer than 1.5 percent of workers in either manager arm mentioned "manager," "person," or "people" in their answer and fewer than 5 percent of workers in the algorithm arm mentioned "algorithm" or "AI." The most common hypothesis was that there were pre-determined answers to compare to. While proofreading, workers did not know they might interact with their manager again.

men to behave differently in the workplace (e.g. Niederle and Vesterlund, 2007; Coffman, 2014; Coffman et al., 2021). Racial minority and white workers may also react differently when they perceive discrimination, since racial minorities experience more systematic discrimination in the workplace (Kline et al., 2021). This familiarity may make the experience either more *or* less upsetting for racial minority workers than white workers.

In my sample, race and gender are highly correlated in the expected ways with past experiences of discrimination, beliefs about the prevalence of discrimination, and confidence and competitiveness in the science, grammar, and spelling quizzes collected in the baseline survey (Appendix Table A7). Race and gender are more weakly correlated with several other demographic variables like education, income, and employment outside Prolific, which could also affect the cognitive or emotional effects of perceived discrimination. For example, workers with better outside options may be less invested in my job and therefore less affected.

This section of the analysis proceeds in two parts. First, I show there is substantial racial and gender heterogeneity in the effects of being in the non-blind manager arm on retention and performance. This documents the important fact that perceived discrimination exacerbates gender and racial gaps. Then, I show that the other variables related to the potential psychological mechanisms seem to predict these heterogeneous effects by race and gender as expected according to the framework, but cannot fully explain the race and gender differences. Since my analysis sample is made up of white and racial minority women and racial minority men, I focus on racial differences among women and gender differences among racial minorities.

Heterogeneous treatment effects. There are large negative effects on retention for both white and racial-minority women. In the required paragraphs where selection cannot play a role, racial-minority women seem to put in more effort, but only somewhat successfully. In contrast, racial-minority men experience positive performance effects. Figure 8 plots the effect of being in the non-blind manager arm on the cumulative percent quit, bonus earned, and correct and incorrect highlights over all eighteen paragraphs, separately by race and gender. I estimate heterogeneous treatment effects by interacting T_i in Equation 1 with race $\times$ gender group indicators.

Retention. Being in the non-blind manager arm causes white women to proofread 0.65 fewer paragraphs (4 percent, se=0.32), and they are 4pp (70 percent, se=2.2pp, p=0.07) more likely to quit as soon as they are able. Racial-minority women proofread 0.95 fewer paragraphs (6 percent, se=0.55, p=0.05) and are 12pp (15 percent, se=5.7pp) less likely to proofread all of the paragraphs. The effects are not significantly different for white versus racial-minority women. In contrast, racial-minority men proofread 0.89 *more* paragraphs, though the estimate is not statistically significant (se=0.56). The effects on retention for men and women are statistically significantly different from each other with p<0.02, including when restricting to racial minorities (Appendix Table A8).

Performance. Racial-minority men and women's performance is affected, while white women's

is not. Being in the non-blind manager arm causes minority men to *correctly* highlight 1.5 (12 percent, se=0.48) more mistakes in the required paragraphs (Appendix Table A9). They appear to be more focused, as being in the non-blind manager arm also causes them to click 4.8 fewer times (14 percent, se=2.6, p=0.07) while highlighting mistakes. On the other hand, minority women highlight about 1.7 more words overall (10 percent, se=0.6), but they are equally likely to be correct or incorrect (separately, these effects are only marginally significant). The effects on correct and incorrect highlights for racial-minority women are marginally significantly different from the null effects for white women (p=0.12). The positive effects on minority women's correct highlights are not significantly different from the positive effects for minority men, but the effects on incorrect highlights are marginally significantly different from the null effects for minority men (p=0.16).[33]

Consistent with these patterns and the hypothesis that they are driven by psychological mechanisms, there is a positive effect of being in the non-blind manager arm on racial minority men's psychological well-being—they are less upset, anxious, and discouraged, in particular. Racial minority women are less motivated, though also less anxious. There is no effect on white women's psychological well-being (Appendix Figure A5).[34]

Mechanisms. As described above, there are many potential psychological mechanisms that could explain this racial and gender heterogeneity; I find that other variables related to these mechanisms do (noisily) predict treatment effects in the expected direction; they cannot, however, explain all of the observed racial and gender heterogeneity.

First, none of the potential mechanism variables independently predict heterogeneous treatment effects as robustly as race and gender itself.[35] This is unsurprising, as many of the mechanisms are expected to reinforce each other in generating gender heterogeneity, in particular. Appendix Figures A6 and A7 show the effects of being in the non-blind manager arm when the effect can vary by each potential mechanism. Many estimates are consistent with the predictions of the organizing framework—for example, point estimates suggest a more negative effect on retention among under-confident workers than over-confident workers—but the effects on retention are only statistically significantly different from each other when considering gender and race. The positive

[33]This plays out as one would expect in terms of total earnings. There is a negative but insignificant effect on total earnings for women and a positive effect for men (Figure 8). If the bonus was a piece-rate per paragraph, the negative retention effect would correspond to the same reduction in total earnings for women. Instead, earnings depended on passing a performance threshold in each paragraph, so the change in retention may not correspond to the effect on total earnings if there is an effect on performance among those who do not quit. Indeed, the ameliorated negative effect on women's total earnings suggests that there is some positive effect on performance among the women who stay, though this can not be quantitatively separated from selection effects.

[34]There are no overall or sub-group effects on self-efficacy or job satisfaction (Appendix Tables A19 and A20).

[35]The variables I observe in this sample of workers are confidence in the baseline quizzes, beliefs about the prevalence of discrimination and experiencing discrimination at work in the past, and proxies for outside options: income, employment, and having a college degree. Unfortunately, I only observe competitiveness in the hiring experiment, so I document that there are gender differences in competitiveness in that sample but cannot test whether that explains the gender heterogeneity observed in effects on retention and performance.

effect on performance is driven by the over-confident and employed (who are more likely men).

To account for the fact that all of these variables are correlated, I next estimate regressions that sequentially add interactions of all of the characteristics of interest with an indicator for being in the non-blind treatment arm, though this is under-powered. While the point estimates on the interactions of the mechanism variables with the indicator for being in the non-blind manager arm continue to predict heterogeneous treatment effects in expected ways, the statistically significant race and gender heterogeneity in the effects remains after accounting for all potential mechanisms (Appendix Tables A10, A11, and A12).

Robustness. Regardless of *why*, the key result is that race and gender seem to systematically predict differential treatment effects on effort and performance. This result is robust to accounting for multiple hypothesis testing concerns around the number of dimensions of heterogeneity tested.[36] Appendix Table A13 shows the results of the procedure described in List et al. (2019), where I account for the four primary outcomes (including future labor supply, discussed next) and the three race $\times$ gender groups considered. The positive effect on performance for racial-minority men maintains a p-value less than 0.04, but the effects on retention for white and racial-minority women separately do not maintain conventional significance levels. The gender heterogeneity results are stronger when pooling all women. The p-value testing whether the effects on the number of paragraphs proofread is the same for men and women is 0.008 (Appendix Table A8), which maintains significance at the 10 percent level when multiplied by 10—the number of heterogeneity groups in Appendix Figures A6 and A7—as a conservative Bonferroni adjustment.

In summary, perceived discrimination affects worker retention and performance but these results vary significantly by race and gender. Some workers may react by trying to prove themselves to biased managers, but others disengage. Even if future evaluations are race- and gender-neutral, perceived discrimination may therefore affect the outcomes of future evaluation via workers' prior performance. The results suggest that perceived discrimination may exacerbate racial gaps among women and gender gaps conditional on race.

3.4 Effects of perceived manager discrimination on future labor supply

Perceived discrimination reduces future labor supply. These effects can be partially explained by beliefs about lower expected wages but derive primarily from disutility from additional interactions with a biased manager or employer.[37]

[36] The robustness of the overall results to concerns around the number of *outcomes* is discussed in Section 6.

[37] The primary analysis of these outcomes comes from the promotion experiment so that the sample is the same as in the previous section. A subset of these outcomes were also measured in the hiring experiment; this was pre-

3.4.1 Effects of being in the non-blind manager arm

After workers complete the proofreading job, I remind them of how they were assigned, explain that some workers will be offered additional work, and elicit reservation wages for these opportunities to be promoted or not promoted using an incentivized multiple price list (MPL). The reservation wages are measured in terms of both the non-promoted piece rate *and* the promoted piece rate simultaneously, with the higher piece rate always twice the lower. In other words, each row of the MPL reflected two wages, e.g. "Earn \$0.20 per paragraph if not promoted and \$0.40 per paragraph if promoted."[38] Table 4 shows effects on reservation wages calculated as if all workers assume the *same expected wages* in each row, i.e. expected the same probability of promotion.[39] I use a 50 percent probability because the average worker in the control group believes this is the likelihood of future promotion.[40] Thus, an effect on this formulation of reservation wages could be driven either by beliefs about future promotion probabilities or psychological mechanisms like anticipated disutility. I decompose these channels in Section 3.4.2.

Perceived discrimination increases reservation wages for future work and willingness to pay for a transparent, unbiased mechanism. Being in the non-blind manager arm increases reservation wages from 26.5 cents to 28.8 cents per paragraph (2.3 cents, 9 percent, se=1.1) if workers would be evaluated by the same manager (Table 4). Those in the non-blind manager arm are also willing to pay an additional 1.2 cents per paragraph (se=0.5) to be evaluated by a cutoff rule in baseline performance in which there is no potential for future discrimination or interaction with their manager. This willingness to pay, measured as the difference in reservation wages if they would be evaluated by their manager or the cutoff rule, is economically meaningful: it is 4.5 percent of the average reservation wage if one anticipates not being promoted in the next round.

Perceived discrimination also reduces workers' willingness to interact with and generosity towards their manager. Being in the non-blind manager arm increases workers' willingness to pay to be able to choose a different manager for the cooperative job by 5 cents per paragraph (23 percent, se=2.1), though it doesn't affect their reservation wage to do the job without an option to change

registered as a replication exercise and the results are in the *Online Supplement*. The results are generally similar, and described in the text when not.

[38] Workers were told that one row of the MPL would be randomly selected to determine the wages for future work opportunities, and that their answer in that row would determine whether they would be evaluated in the future and offered their assigned (promotion or non-promotion) job at the given wage. They were told that if interest in the opportunity at a given wage exceeded demand, workers who were interested in the jobs at the randomly chosen wage schedule would be randomly selected. 200 workers were evaluated and offered one of the two jobs. The full elicitations for all MPLs in the experiment are in the *Online Supplement.*

[39] I follow convention and drop observations in which workers make inconsistent choices indicating they do not understand the question. Effects on retention and performance are similar in this sample (Appendix B).

[40] This puts this overall reservation wage on the same scale as the variable that accounts for workers' beliefs about future promotion in the next section. 50 percent is much higher than the actual promotion rate (which workers had no information about); a large share of workers guess 50 percent but there is substantial variation.

managers (Table 5). Workers in the non-blind manager arm are also 4.9pp (9 percent, se=3pp, p=0.10) more likely to share \$0 with their manager of a possible \$20 bonus, and share about 29 cents less with their manager, though this is not statistically significant (Table 5).[41]

Applied to a work context where employees and managers interact on a regular basis, this response to perceived discrimination could be self-fulfilling—if workers react to perceived discrimination by avoiding their managers or refusing their requests, managers may be less likely to reward these workers in the future even if they are unbiased. If workers subsequently express anger at work, the effects may be even more damaging (Porat and Levy Paluck, 2024). If managers are biased, these effects could even further compound the effects documented by Glover et al. (2017), who find that workers exert less effort for biased managers because these managers pay them less attention. Biased managers may also react more negatively to women or racial minorities who are less cooperative or generous. Overall, both the reduction in future labor supply and potential follow-on effects to worker retaliation suggest that perceived discrimination would exacerbate gender and racial gaps in labor supply and promotion relative to white men.

3.4.2 Mechanisms

The effects on willingness to interact with and generosity towards managers, as well as the effects on retention when no further discrimination is expected, suggest that animus or avoidance may underly the effects of perceived discrimination on worker behavior. The effects on reservation wages, however, could also be attributed to higher beliefs about the likelihood of future discrimination and thus lower expected wages. Past work on anticipated discrimination either focuses on the second channel—lower returns to effort—or does not separate the two.[42]

Understanding how each channel contributes to the effects of perceived discrimination sheds light on whether models of anticipated discrimination that focus on changing returns to effort may be missing a key, psychologically-driven element. Furthermore, large psychological costs would further amplify the disparate, harmful impacts of a discriminatory labor market on minority-group workers relative to an analysis that only considered the material costs.

[41] This "dictator game" measures distributional and retaliatory preferences in this setting. Workers whose managers did not see demographic information shared on average \$2.81 (14 percent of the total) with their manager. A meta-analysis suggests that workers would have shared around \$6 (30 percent) with strangers (Engel, 2011), so even workers in the demographic-blind manager arm may be retaliating against their managers for not promoting them, or the hierarchical nature of the manager-employee framing may affect distributional preferences. The effect of being in the non-blind manager arm is on top of any effect of non-promotion or the manager-employee framing.

[42] Theory focuses on returns to effort and lab experiments testing these theories randomly assign workers to the discriminated "identity," potentially making psychological costs less relevant (Lundberg and Startz, 1983; Coate and Loury, 1993; Fryer et al., 2005; de Haan et al., 2017; Dianat et al., 2020; Campos-Mercade and Mengel, 2024). Experiments studying anticipated identity-based (gender, age, sexual orientation, etc.) discrimination do not separate (incorrect) beliefs about discrimination from avoidance of psychological costs (Alston, 2019; Charness et al., 2020; Lepage et al., 2022; Ridley, 2022; Agüero et al., 2023; Aksoy et al., 2023; Avery et al., 2023; Angeli et al., 2024).

Many papers have linked perceived discrimination to worse psychological well-being and physical health, with stress as a potential moderator (e.g. Kessler et al., 1999; Pascoe and Smart Richman, 2009; Thoits, 2010; Schmitt et al., 2014), and perceived discrimination at work to lower job satisfaction, organizational citizenship behavior, and employer loyalty, and higher job search intentions (Ensher et al., 2001; Stainback and Irvin, 2012; Hahn and Wilkins, 2013; Mukerjee, 2014; Triana et al., 2015). Black workers who report previously experiencing discrimination search for jobs more widely (Pager and Pedulla, 2015) and controlling for perceived discrimination closes black-white gaps in labor supply (Goldsmith et al., 2004).

Simple framework. My experimental design disentangles the role of material versus psychological costs in explaining the documented effects of perceived discrimination on future labor supply by mapping onto a simple framework. In a standard model, perceived discrimination affects reservation wages only through beliefs about the probability of future promotion. Reservation wages equalize expected utility from work and an outside option, where

$$\text{Expected utility} = \hat{\pi}U(w_H) + (1-\hat{\pi})U(w_L),$$

$\hat{\pi}$ is beliefs about the likelihood of future promotion, w_H is the higher wage if you are promoted, w_L is the lower wage if you are not, and $U(w)$ is an indirect utility function representing utility at a given wage w when a worker chooses optimal effort. Utility is weakly increasing in the wage.

Let D indicate perceived discrimination in a previous encounter and $\hat{\sigma}$ be beliefs about the likelihood of future discrimination. In the standard model, perceived discrimination can only lower expected utility and thus raise reservation wages if beliefs about future discrimination are increasing in perceived discrimination: $\hat{\sigma} = \sigma(D)$, with $\frac{\partial \sigma}{\partial D} > 0$, and if beliefs about future promotion are a decreasing function of anticipated discrimination: $\hat{\pi} = \pi(\hat{\sigma})$, with $\frac{\partial \pi}{\partial \hat{\sigma}} < 0$. This simple framework has two implications: (i) when future discrimination is impossible, perceived discrimination should not affect beliefs about promotion, and (ii) after taking into account any effects on beliefs about future promotion, there should be no effect of perceived discrimination on reservation wages.

In a richer framework, perceived discrimination could affect expected utility and thus reservation wages through two other channels. First, beliefs about future promotion may depend on D independently of $\hat{\sigma}$: $\hat{\pi} = \pi(\hat{\sigma}, D)$; for example, because perceived discrimination changes self-confidence or trust in the employer. Second, workers may experience or anticipate a cost $C(\hat{\sigma}, D)$ which can affect expected utility independently of $\hat{\pi}$: workers may dislike discrimination or interacting with a biased manager or an employer who hires biased managers, conditional on some expected promotion probability. The anticipation cost is increasing in $\hat{\sigma}$ and D. Altogether, now:

$$\text{Expected utility} = \pi(\hat{\sigma}, D)U(w_H) + (1-\pi(\hat{\sigma}, D))U(w_L) - C(\hat{\sigma}, D).$$

Results. The evidence from my setting rejects both implications of the standard model. First, anticipated discrimination explains about half of the effect of perceived discrimination on beliefs about the likelihood of future promotion, suggesting that there are other changes in self-confidence or employer trust, for example. Second, the resulting change in expected utility from wages explains only 30 percent of the overall effect on reservation wages, though this estimate is noisy. The effect on disutility $C(\hat{\sigma}, D)$ explains the rest.

First, I estimate effects on beliefs about the likelihood of future promotion, separately when future discrimination is possible (evaluation by the same manager, i.e. $\hat{\sigma} \geq 0$) and when it is not (promotion based on a cutoff rule in baseline performance, i.e. $\hat{\sigma} = 0$). In the standard model there may be an effect when $\hat{\sigma} \geq 0$, but not when $\hat{\sigma} = 0$.

Being in the non-blind manager arm increases anticipated discrimination *and* affects beliefs about promotion when future discrimination is impossible, though only the former is significant. Being in the non-blind manager arm lowers beliefs about the likelihood of future promotion by 3.2pp (7 percent, se=1.2pp) when workers would be evaluated by the same manager. About 60 percent of this effect remains after accounting for changes in beliefs about the likelihood of promotion under the cutoff rule. 2pp (se=0.8pp) of the effect of perceived discrimination on beliefs about the likelihood of future promotion by the same manager therefore comes from increased anticipated discrimination (Table 6).

Next, I estimate the effects of perceived discrimination on reservation wages, constructed to account for changes in beliefs about the likelihood of promotion. Any effect on this construction of reservation wages isolates the role of $C(\hat{\sigma}, D)$. Recall that the reservation wages described earlier (Table 4) were constructed from the MPL in which each row had both a non-promotion and promotion wage, assuming that workers had the *average* worker's belief about promotion probabilities. In this case, movement along the MPL can reflect different valuations of the job *or* changes in beliefs about the probability of promotion. To isolate changes in valuations of the job (psychological costs, in my framework), the high and low wages in the MPL are multiplied by workers' *own* belief about the probability that they would be assigned to each type of job. Average reservation wages in the demographic-blind manager arm using both measures are the same and the treatment effects can be interpreted on the same scale. Effects on reservation wages constructed in this way when future discrimination is not possible isolate $\frac{\partial C}{\partial D}$, or distaste for working for an employer who previously hired biased managers; when future discrimination is possible, effects can additionally stem from $\frac{\partial C}{\partial \hat{\sigma}}$, or disutility from anticipating additional discrimination.

Being in the non-blind manager arm increases reservation wages for future evaluation by the same manager (i.e. when future discrimination is possible) by 1.6 cents per paragraph (a 6 percent increase, se=1 cent, p=0.12) using this construction of reservation wages (Figure 9, top panel), suggesting that psychological costs are non-zero. Taking the estimated coefficients at face value,

70 percent of the total effect on reservation wages (2.3 cents per paragraph) is due to psychological costs, however, there is substantial overlap in the confidence intervals of the two point estimates.

When workers will be evaluated by the cutoff rule, treatment increases reservation wages by 0.7 cents per paragraph after accounting for beliefs (though this is statistically insignificant). Again taking the point estimates at face value, this means that 44 percent of the overall effect on $C(D, \sigma)$ is due to disutility from anticipated discrimination or continued interactions, and the remainder is due to disutility from continued work for the employer in general. Consistently, being in the non-blind manager arm reduces whether workers strongly agree that they are interested in future work for the employer by 5.8pp (9 percent, se=2.9pp); they are also less interested in tasks with similar assignment procedures, or doing the easy or hard paragraphs in the future, though these point estimates are not significantly different from zero (Appendix Table A14).

Note that the future work opportunities studied here shut down reasons to avoid future work driven by direct interactions with the manager in the future—for example, micro-aggressions in day-to-day interactions—suggesting that the change in reservation wages and role of psychological costs documented here may be a lower bound. The large effects on generosity towards managers and willingness to pay to choose a new manager for a cooperative task described earlier are consistent with this, but cannot be quantitatively benchmarked to these results.

Altogether, these results suggest that the standard model cannot explain the effects of perceived discrimination on future labor supply. This is consistent with the effects on retention and performance which cannot be due to anticipated discrimination. Thus, models of anticipated discrimination that focus on changing returns to effort may be missing an important, psychologically-driven element. Discrimination may have even more pernicious effects than previously considered because it has a direct utility cost. Even if workers face no monetary consequences, perceived or anticipated discrimination makes them worse off.

3.4.3 Heterogeneous treatment effects

There is not obvious racial or gender heterogeneity in the effects of perceived discrimination on future labor supply. Being in the non-blind manager arm increases all workers' willingness to pay to avoid their manager and causes retaliation, though minority men are more retaliatory and white and racial minority women are more avoidant (Appendix Table A15). This suggests that the effects of perceived discrimination on labor supply would exacerbate gender and racial gaps across the board for under-represented groups in STEM, relative to their white male counterparts.

Certain workers may still experience larger psychological costs of perceived discrimination or expect more discrimination (even after accounting for their stated beliefs about the future likelihood of promotion). These are workers who think discrimination is more prevalent in general or

report experiencing discrimination at work in the past. Workers with better outside options may also be more avoidant or retaliatory if they are less reliant on their income from Prolific.

The proxies I observe for all of these characteristics do not, however, predict significant heterogeneity in the effect of being in the non-blind manager arm on workers' reservation wages, willingness to pay for an unbiased mechanism, cooperative task reservation wage, or willingness to pay for a new manager. Point estimates suggest that *all* workers want to avoid future interactions with biased managers. However, some workers are more retaliatory than others in expected ways. Those who have above-median beliefs about the prevalence of discrimination in general, who report experiencing discrimination in the past, who are over-confident, and who have a college degree are much more retaliatory (Appendix Figures A8 and A9 show results for each characteristic individually; results accounting for them all simultaneously are similar and available upon request).

Together, the results so far show that perceived discrimination has quantitatively important effects on worker behavior and subsequent outcomes. These effects would exacerbate gender and racial gaps. Addressing the disparate impacts of discrimination therefore requires addressing worker perceptions alongside discrimination itself. In addition to these equity concerns, these results show that employers face an efficiency rationale to reduce perceived discrimination or mitigate its effects. They could improve retention and worker-manager cooperation by doing so.

3.5 Effects of perceived algorithmic discrimination

Hiring and promotion decisions are increasingly based on algorithmic screening (Cowgill, 2020; Jarrahi et al., 2021). Workers do perceive algorithmic discrimination in the experiment—even when they know the algorithm did not have access to race and gender (Section 3.2)—so I can test whether workers' reactions to perceived algorithmic discrimination are similar to their reactions to perceived manager discrimination. People are less morally outraged by depictions of algorithmic bias than human bias and are less likely to attribute algorithmic bias to prejudice (Bigman et al., 2023); if the effects documented so far are due to perceived *prejudice* rather than perceived *discrimination,* they may not arise when workers perceive algorithmic discrimination. On the other hand, if perceived discrimination affects workers through a stress response or attention capture, there may still be negative effects. The effects of perceived algorithmic discrimination seem to be negative, but differ from the effects of perceived manager discrimination in ways that are consistent with these contrasting psychological drivers.

3.5.1 Retention and performance in the proofreading task

Learning that the demographic-blind algorithm previously promoted mostly white men has, if anything, a positive effect on retention. The positive effects on the number of paragraphs proofread and whether workers proofread any optional paragraphs are statistically significantly different from the negative effects of being in the non-blind manager arm with $p<0.07$ (Figure 10, Row 1).

On the other hand, learning that the demographic-blind algorithm previously promoted mostly white men negatively affects performance in the required paragraphs. Learning that the demographic-blind algorithm previously promoted mostly white men causes workers to correctly highlight 0.67 fewer words (4.5 percent, se=0.4, p=0.09) and earn 6.2 cents less (7 percent, se=3.7 cents, p=0.09). These effects are statistically significantly different from the effect of being in the non-blind manager arm with $p<0.05$ (Figure 10, Row 3).

3.5.2 Future labor supply

Learning that the demographic-blind algorithm previously promoted mostly white men has no significant effect on reservation wages or willingness to pay for the unbiased cutoff, but the point estimates are positive and not statistically distinguishable from the positive effects in the manager arms (Figure 10, Row 4). These smaller, less distinguishable effects may be due to the fact that there is a smaller, insignificant effect of learning that the demographic-blind algorithm previously promoted white men on beliefs about future promotion or anticipated discrimination (Figure 10, Row 5). In the hiring experiment, however, perceived algorithmic discrimination did significantly increase anticipated discrimination and reservation wages for future work (as did perceived manager discrimination). In this case, there was no role for beliefs about the likelihood of being hired to affect reservation wages and the entire effect would be due to psychological costs. Thus, the anticipated psychological costs of algorithmic and manager discrimination may be similar.[43]

In summary, the use of algorithmic hiring procedures changes the effects of perceived discrimination but the effects are still negative. These results are relatively imprecisely estimated (the sample is half the size, and random assignment generates less of a difference in perceived discrimination rates). Taken at face value, however, they are consistent with the idea that workers react differently to perceived *prejudice* than they do to perceived *discrimination* (Bigman et al., 2023) but that perceived discrimination absent prejudice can still trigger stress responses or attention capture: workers do not disengage in the same way during the task, but experience negative effects

[43] These beliefs and reservation wages were the only outcomes collected in the hiring experiment that related to worker behavior. The *Online Supplement* describes these results, which were pre-registered as a replication exercise. The effects of perceived algorithmic discrimination are much better powered in the hiring experiment.

on their performance and future labor supply. That said, many other things are different about perceived human versus algorithmic discrimination; other potential explanations are differences in *who* perceives discrimination and heterogeneous treatment effects, the fact that algorithms cannot themselves "benefit" from workers' effort, or the fact that an algorithm cannot be "proven wrong."

Overall, perceived manager and algorithmic discrimination negatively affect workers and firms face incentives to reduce perceived discrimination rates. The next two sections test how they might achieve this by changing their hiring or promotion policies.

4 Effects of minority-group representation on perceptions

A key input to workers' beliefs about experiencing discrimination in my context is their reaction to the information that primarily white men were promoted in the past. While there are many reasons that firms may want to increase diversity in their organizations (e.g. Niederle et al., 2012; Gompers and Kovvali, 2018), one under-explored potential benefit is reducing perceived discrimination and its subsequent effects. If greater minority-group representation signals a less-biased manager or algorithm, workers should perceive less discrimination when they see more representation.

The promotion experiment allows me to test how workers' beliefs about discrimination change when they see that one, rather than zero, of the three previously-promoted workers is not a white man. I do this holding fixed the information that workers have about the promotion procedure—in other words, only the outputs of the procedures change, not the procedures themselves.

4.1 Estimation

I estimate the effects of minority-group representation by comparing workers who saw that three white men were previously promoted to those who saw two white men and someone else. There is truthful, random variation in the demographic composition of the previously-promoted workers that each worker saw due to the random assignment of workers to managers and to groups of historical workers that were jointly evaluated by the algorithm (see Figure 3).

In the manager arms, workers were randomly assigned to managers, who made different (true) choices among historical workers. In the non-blind manager arm, 45 percent of workers were randomly assigned to managers who had previously promoted three white men and 55 percent were assigned to managers who had promoted two white men and someone else.

In the algorithm arm, workers were randomly assigned to groups of historical workers who had been jointly evaluated by the algorithm; in some groups three white men were promoted and

in others it was two white men and someone else. When workers learned previously-promoted workers' demographic information, 33 percent saw three white men and 67 percent saw two white men and someone else.[44] Workers who see that two versus three white men were previously promoted are observably similar (Appendix Table A16). I estimate the following:

$$D_i = \alpha + \delta_1 3WM_i + \gamma_1 T_i^M + \beta_1 T^M \times 3WM_i + \theta_1 A_i + \delta_2 A \times 3WM_i + \gamma_2 T_i^A + \beta_2 T^A \times 3WM_i + X'\delta + H'\lambda + \gamma_g + \varepsilon_{ig},$$

where $3WM_i$ is an indicator for being randomly assigned to a group in which three white men were previously promoted, T_i^M is an indicator for being evaluated by a non-blind manager, A_i is an indicator for being evaluated by the demographic-blind algorithm and not learning previously-promoted workers' demographics, and T_i^A is an indicator for being evaluated by the demographic-blind algorithm and learning previously-promoted workers' demographics. D_i indicates perceiving discrimination. All other variables are the same as previously.

The effect of seeing that only white men were previously promoted is β_1 when workers were evaluated by a non-blind manager and β_2 when workers were evaluated by a demographic-blind algorithm. These are the primary coefficients of interest. δ_1 and δ_2 should both be zero, as perceived discrimination rates among workers who did not see previously-promoted workers' demographics should not depend on the demographics of those workers.

As in all previous analysis, the sample is restricted to workers who would not have been promoted under any procedure, and standard errors are robust to heteroskedasticity. This analysis continues to drop the small sample of white men, but the results are unchanged if white men are included.[45] Note that this analysis was not pre-registered (see Section 6).

4.2 Results

More representation of women and racial-minority men among previously-promoted workers substantially reduces perceived discrimination, even when managers or algorithms could have discriminated. Seeing one minority-group member among the three previously-promoted workers reduces rates of perceived discrimination by 19pp (41 percent, se=4pp) in the non-blind manager arm and by 15pp (48 percent, se=5pp) in the algorithm arm when workers see previously-promoted workers' demographics (Figure 11).[46]

[44] This split is different than in the manager arms because the algorithm was less likely to hire three white men—the key is that *within* each arm, whether workers saw two or three white men being promoted was random.

[45] White men may perceive "reverse racism" when more minority-group members are hired (Norton and Sommers, 2011), but my data does not provide an opportunity to test this because the sample of white male workers is so small.

[46] This additional dimension of random variation in perceived discrimination provides another avenue to test the effects of perceived discrimination. This analysis is underpowered and not pre-registered, but generally consistent: seeing

Whether the previously-promoted minority-group worker is of the worker's same demographic group also matters. I add indicators for seeing one previously-promoted worker from a workers' own demographic group and the corresponding interactions to the above regression. In the non-blind manager arm, seeing one previously-promoted minority-group worker from a *different* demographic group as oneself reduces perceived discrimination from 44 to 34 percent relative to seeing three white men (a 10pp difference, se=4pp). This is consistent with theory and evidence for "stigma transfer:" when presented with information that an actor is prejudiced against one demographic group, it is common for other stigmatized groups to feel threatened (Sanchez et al., 2017, 2018). In my case, evidence that an actor is *not* prejudiced against a given group also seems to transfer across groups. That said, seeing one minority-group member from one's *own group* has twice the effect: this reduces perceived discrimination by an *additional* 18pp (se=5pp). The same occurs in the algorithm arm (Appendix Figure A10).[47]

These results do *not* suggest that improving minority-group representation by any means necessary is a silver bullet to issues of perceived discrimination. They do suggest that study of other policies concerned with minority-group representation—e.g affirmative action—should consider the channel of perceived discrimination and perceptions of the policy more broadly. Reducing initial perceived discrimination (e.g. in hiring) without reducing actual discrimination (e.g. in day-to-day activities) would likely backfire. Majority-group backlash to these types of policies may exacerbate this issue (Norton and Sommers, 2011; Jardina, 2019; Leslie, 2019) and should be measured as well. This relates to why initiatives to improve diversity and inclusion (including employee buy-in) occur in parallel (e.g. Hrala, 2022).

5 The effects of anti-bias policies on perceptions

Changing the outcomes of existing hiring and promotion procedures is likely to be difficult in settings where minority-groups are highly under-represented, as explicit affirmative action and related diversity initiatives are increasingly contentious (e.g. Sorkin et al., 2023). Firms may be able to more easily change what types of information decision-makers have about candidates or how the decision is made (by a manager, algorithm, or in combination).[48] The results in the previous section suggest that if these new procedures hire or promote more diverse candidates,

that three white men were previously promoted increases perceived discrimination and amplifies the magnitude of the effects of being in the non-blind manager arm in the direction of the average effects on worker behavior.

[47] The conclusions are largely similar when estimated separately for each demographic group, though sample sizes per arm become small and results are therefore quite imprecise; they are available upon request. The main difference is quite rational: racial-minority women share some identity with either of the other groups in this context, and seem to be more responsive to seeing members of a different minority group than white women or racial-minority men.

[48] They could also try to de-bias managers explicitly, but this may be difficult (Chang et al., 2019; Paluck et al., 2020).

their implementation is likely to reduce perceived discrimination rates. In many settings with low diversity in *applicant pools*, however, even de-biasing procedures may not lead to more minority-group representation in the short run. Thus, to mitigate the effects of perceived discrimination, it is important to understand how workers perceive discrimination under different hiring procedures if there continues to be minimal minority-group representation.

Firms are also increasingly turning to algorithmic decision-making to reduce hiring costs or improve employee fit, in some cases at the expense of diversity rather than with diversity as a goal. Academic and popular opinion often focuses on the potential for algorithms to "bake-in" human biases, though they also have the potential to reduce the prevalence of discrimination in the long run (e.g. Kleinberg et al., 2018; Obermeyer et al., 2019; Kleinberg et al., 2020; Raghavan et al., 2020). Little is known, however, about how workers *perceive* algorithmic bias, specifically.[49]

5.1 Estimation

The hiring experiment cross-randomized workers to be evaluated by a manager or algorithm who either knew demographics or not, and further cross-randomized workers evaluated by a demographic-blind decision-maker to either learn the demographics of previously-promoted workers or not. I estimate the following:

$$D_i = \alpha + \eta Alg_i + \theta_1 BA_i + \theta_2 NB_i + \chi_1 BA \times Alg_i + \chi_2 NB \times Alg_i + X'\delta + H'\lambda + \gamma_g + \varepsilon_{ig},$$

where Alg_i is an indicator for being randomly assigned to be evaluated by an algorithm rather than a manager, BA_i is an indicator for being randomly assigned to be evaluated by a demographic-blind decision-maker and to see previous hires' avatars, race, and gender, and NB_i is an indicator for being evaluated by a non-blind decision-maker and seeing previous hires' avatars, race, and gender. D_i indicates perceiving discrimination. All other variables are the same as previously.

The effect of being evaluated by an algorithm is η when decision-makers are demographic-blind and workers do not know previous hires' race and gender, and χ_1 (χ_2) when the decision-maker is demographic-blind (non-blind) and workers know that previous hires were mostly white men. The effect of learning that a demographic-blind manager previously hired mostly white men is θ_1 and the effect of learning that a manager knew race and gender is $\theta_2 - \theta_1$. The effect of learning that a demographic-blind algorithm previously hired mostly white men is $\theta_1 + \chi_1$ and the effect of learning that an algorithm used race and gender is $\theta_2 - \theta_1 + \chi_2 - \chi_1$.

The hiring experiment specifically tests whether introducing unbiased evaluation procedures is sufficient to reduce perceptions of discrimination *when minority-group workers still do not see*

[49] Whether people judge humans or algorithms as more "fair" depends on the definition of fairness (Lee, 2018; Kaibel et al., 2019; Acikgoz et al., 2020; Newman et al., 2020; Noble et al., 2021; Zhang and Yencha, 2022).

themselves represented among previously-promoted workers. That is, the unbiased procedures have not led to any observable changes in decision-making—a realistic outcome when minority-groups are severely under-represented. In the hiring experiment, almost all workers saw exclusively white men being hired by design. In the main analysis, the above regression is estimated only on those workers (a random subset, as in the promotion experiment).[50]

Like the promotion experiment, the hiring experiment facilitates the counterfactual evaluation of each worker by every procedure and the sample restricts to workers who would not have been hired regardless of their random assignment (1,934 of 2,054 in the subset seeing all white men). Results are similar with the full sample. Standard errors are robust to heteroskedasticity. The results do not depend on any of these specification decisions (Section 6).

5.2 Results

Changing hiring procedures in a way that reduces or eliminates the scope for human bias to affect decisions cannot eliminate perceptions of discrimination when workers see that only white men were previously promoted. Figure 12 plots rates of perceived discrimination in each treatment arm in the hiring experiment estimated using the main specification. No workers perceive discrimination in the demographic-blind manager or algorithm arms where workers do not see previously-promoted workers' demographics, so I focus on the other arms.

Workers perceive *more* discrimination by algorithms than managers. When decision-makers see workers' avatars, race, and gender and workers know all previous hires were white men, 52 percent of workers evaluated by a manager perceive discrimination, and workers evaluated by an algorithm are 10pp (se=5pp) more likely to perceive discrimination. This difference is similar when decision-makers do not know demographics (a 9pp difference, se=4pp).

Blinding decision-makers to demographics reduces perceived discrimination, but not nearly to zero when workers know that all previous hires were white men. Specifically, blinding decision-makers reduces perceived discrimination by 11pp (21 percent, se=5pp) and 12pp (19 percent, se=3pp) when workers are evaluated by managers and algorithms, respectively, but 42 percent and 51 percent of workers still perceive discrimination.

The directional effects of algorithmic decision-making and blinding decision-makers to demographics are similar, though not always statistically significant, using the secondary measures of perceived discrimination (Appendix Figure A11). Like in the promotion experiment, the levels of these secondary measures vary but are correlated with the main measure.

[50] Results for the full sample are similar but conflate perceptions of procedures with minority-group representation. (Appendix Figure A17). The procedures are differentially likely to hire entirely white men in the historical sample; this affects perceived discrimination rates in the same ways as observed in the promotion experiment.

The high rates of perceived discrimination that persist when decision-makers are demographic-blind are due to workers not attending to or believing this information. While 49-58 percent of workers in the other four arms correctly identify the information that their manager or the algorithm knew when making hiring decisions, only 6 (16) percent of workers do so when they are evaluated by a demographic-blind manager (algorithm) if they see previous hires' avatars, race, and gender. They are more likely to be incorrect precisely because they think the decision-maker knew their race and gender. 95-97 percent of workers correctly identify that the non-blind manager and algorithm knew race and gender, but 87 (70) percent of workers *incorrectly* think so when they are evaluated by a demographic-blind manager (algorithm) but see that the previously-promoted workers are all white men (Appendix Figure A12). This does *not* seem to be due to differential inattention or general confusion; Section 2.4 shows that attention and all other comprehension is high and similar across treatment arms.

Potential confounds. That said, I cannot separate disbelief or motivated reasoning from differential forgetfulness or confusion about *this particular* decision input. The information about decision inputs and who was previously promoted were communicated similarly, with workers shown their own evaluated profile separately from the profiles of the previously-promoted workers. This was meant to help workers attend equally to the decision inputs and outputs, but may not have been successful. Workers may continue to perceive discrimination when managers are blinded because workers do not realize they are blinded, which may not be generalizable to other settings. An alternative is that workers do not *believe* that the decision-makers were blinded, implying that trust in employers may be an important mediator.

Potential psychological mechanisms. Despite potential confounds, there are also two well-documented psychological mechanisms that may help explain workers' belief that the managers were biased even when they are told that the manager did not know their race or gender: the representativeness heuristic and attribution bias. The representativeness heuristic is the predisposition to assume that events that are "representative" are also probabilistically likely (e.g. Kahneman and Tversky, 1972; Bordalo et al., 2021). In this case, workers have seen a decision that "looks like" discrimination, and thus jump to the inference that discrimination played a role in the decision. Another possible explanation is attribution bias, specifically the fundamental attribution error (e.g. Ross, 1977; Ross and Nisbett, 1991): when evaluating other people's behavior, people put more stock in the role of individual personality traits than the potential role of situational factors. In this case, this would be the attribution of the decision to hire only white men to manager bias rather than thinking of other situational possibilities, like an overrepresentation of white men in the applicant pool. If driven by either psychological mechanism, the result that workers think that blinded managers knew demographic information would be interpreted as an incorrect cognitive encoding of information to make their perception of discrimination plausible and internally consistent.

Taking together the results from Sections 4 and 5, there are several implications for settings where some demographic groups are severely under-represented. First, anti-bias policies—like transitioning to blinded algorithmic screening or censoring demographic identifiers from resumes before manager review—may slightly reduce perceptions of discrimination on their own, but workers may still be likely to perceive that discrimination played a role if there is no change in the demographics of workers who are hired. On the other hand, if they improve minority-group representation, they may be more effective at reducing perceived discrimination. These results show that increasing minority-group representation from 0 of 3 to 1 of 3 cuts rates of perceived discrimination in about half in this setting, but cannot speak to the *degree* of representation that is needed to obtain more favorable perceptions of hiring procedures. Research on reactions to tokenism—when a small number of members of a disadvantaged group are seated in positions of power—suggests that there is likely some such threshold (e.g. King et al., 2010; Anisman-Razin and Saguy, 2016). On the other hand, people tend to overestimate diversity when seeing a few members of minority groups (Kardosh et al., 2022), so representation rates may not need to reach parity.

Taking together the results from Sections 3.5 and 5 yields important considerations regarding the proliferation of algorithmic hiring and promotion procedures. If workers cannot be convinced that algorithms are demographic-blind, their use is likely to *increase* perceived discrimination rates. Since perceived algorithmic discrimination has negative effects on worker performance and future labor supply, the transition to algorithmic decision-making is unlikely to be a solution to disparities caused by perceived discrimination.

6 Robustness

The results in Sections 3, 4, and 5 are broadly robust to accounting for standard concerns with experiments in general and survey experiments in particular. This section provides a broad overview of these analyses; Appendix B provides details and corresponding tables and figures.

Multiple hypothesis testing and randomization inference. The statistical significance of my results does not change when I relax the normality assumptions associated with asymptotic standard errors and instead use treatment randomization to estimate the distribution of treatment effects under a null hypothesis of no treatment effect (Fisher, 1935; Rosenbaum, 2002). The same is true when I account for multiple hypothesis testing given the number of outcome variables using a Romano-Wolf (2005; 2016) correction, with the exception of the negative effect of perceived algorithmic discrimination on worker performance.

Attrition and attention. There is no differential attrition or attention that can explain my

results. Section 2.4 discusses overall experimental fidelity and shows that worker attention and comprehension was high overall. The specification charts described next show that the results are generally robust to dropping workers who failed the attention check.

Specification choices. The appendix includes specification charts that iterate over every combination of specification choices. The effects of perceived manager discrimination and the effects of minority-group representation and anti-bias hiring procedures on perceived discrimination rates are unchanged if I restrict the sample to workers who passed the attention check or expand the sample from workers who would not have been promoted or hired under any procedure to the full sample. The only non-robust result is that the effects of perceived algorithmic discrimination on performance and retention are closer to zero and statistically insignificant when restricting to those who passed the attention check.

All four sets of results are very stable regardless of the control variables used, consistent with effective randomization and lack of observable differences between treatment arms. The results are robust to replacing the specifications with the exact pre-registered specification, using lasso-selected control variables (Chernozhukov et al., 2018), or using no controls.

Experimenter demand effects. The experimental surveys asked workers to indicate what they thought the survey was about at multiple points throughout the survey (timing was random) in an open-ended text response. I use coded measures of whether they think the study is about discrimination (described in the *Online Supplement*) to understand the extent to which experimenter demand effects—differential knowledge of the study purpose and resultant changes in worker behavior—could explain my main results. While there were differences across treatment arms in the share of workers who believed the study was about discrimination, the rates at which workers did so remained low until after the more explicit measures of perceived discrimination were elicited at the very end of the survey. To test whether these differences affected worker behavior in a way that explains my treatment effects, I bound my results by replacing outcome variables with higher or lower values for workers who believed the topic of the study was discrimination. I replace the worker behavior outcomes with plus or minus 0.2sd of the variable in the control group (following the results in de Quidt et al., 2018), and I replace the perceived discrimination outcome with 0 or 1—providing more conservative bounds since this outcome was not incentivized.

The effects of perceived discrimination on worker behavior are unchanged by this exercise.[51] The effects of changing hiring procedures and the effect of minority-group representation on perceived discrimination fall by up to half when replacing the perceived discrimination measure with zero for workers who believe the study is about discrimination, but generally remain significant. The conservative interpretations are that increasing minority-group representation cuts perceived discrimination rates by 25 percent (rather than 45), using an algorithm increases perceived dis-

[51] In prior work, workers do not change their behavior when told a study's purpose (Mummolo and Peterson, 2019).

crimination by about 15 percent (rather than 20), and blinding decision-makers reduces perceived discrimination by 8-10 percent (rather than 20).

Multiple price list elicitations. The four measures of future labor supply were based on multiple price list-style elicitations of reservation wages and willingness to pay in terms of foregone wages. These were incentivized following the literature by telling workers I would randomly choose one row of the list and implement randomly-chosen workers' choices for that row, though the incentives were relatively low-powered as workers could report low reservation wages and then choose to not take the job if offered later. However, only 10-15 percent of workers answered "I would want the job at this wage" for all wages and the results are robust to dropping these workers. In the main results, I follow the literature by dropping any worker who does not respond to the MPL questions in a sensical way; Appendix B describes these workers and shows that the main results are robust to dropping these workers from all analyses.

Deviations from Pre-Analysis Plan. I posted detailed pre-analysis plans before each experiment, and follow them closely. Here, I outline the main deviations; Appendix B describes details and smaller deviations.

The primary deviation was to exploit random variation in the fraction of previously-promoted workers that were white men, which was built into the original experimental design by randomly assigning workers to managers and jointly-evaluated groups of historical workers. I did not anticipate the amount of variation in the composition of previously-promoted workers. Given that (i) there was substantial variation and (ii) the effect of minority-group representation on perceived discrimination is an important comparison to the effects of procedural changes, I include these results prominently in the paper.

There were also two deviations related to recruitment. First, in the promotion experiment I planned to more substantially over-sample racial minorities. I was unable to recruit the pre-registered number of racial minorities on Prolific. That said, I am still able to test for racial and gender heterogeneity and obtain moderately precise results. Second, in both experiments, the sample size of workers offered the experimental survey is 91-93 percent of pre-registered sample size. This is due to difficulty in recruiting managers and in bringing them back to participate in the second and third rounds of worker evaluations.

Finally, in the main analysis I drop the white men in the promotion experiment since the treatment did not generate perceived racial or gender discrimination in this sample (as intended), though this was not pre-registered. The results are unchanged if I include these workers.

7 Conclusion

This paper shows that addressing the disparate impacts of discrimination will require addressing worker perceptions in addition to discrimination itself. Perceived discrimination affects retention, performance, and future labor supply in ways that exacerbate gender and racial gaps caused by discrimination itself. Workers are also willing to pay to avoid discrimination—beyond just its expected economic consequences—implying that a discriminatory labor market has disparate psychological costs as well as material ones.

While employers stand to benefit from higher retention and easier or cheaper recruitment by reducing perceptions of discrimination, I show that it is likely going to be difficult to do so purely by implementing anti-bias procedures in contexts where minority-group workers are severely underrepresented, unless they are very effective at increasing minority-group representation. Reducing biased decision-making itself is of course the right place to start—and changing worker perceptions without changing actual bias is likely unsustainable. But *just* changing procedures so that they have less potential for bias—in this case, blinding decision-makers to race and gender or using unbiased algorithms—at best only mildly reduces perceived discrimination if the outcomes of those procedures remain the same and decisions continue to "look like discrimination." Increased diversity may be a more promising tool, and perceptions of discrimination under affirmative action or other diversity initiatives warrant future study. In addition, when thinking about a standard workplace, one must consider perceptions of bias in other interactions and procedures in tandem with changing actual or perceived bias in hiring or promotion practices.

These results also offer other interesting avenues for future research on perceptions of discrimination. I studied particular anti-bias tools and communicated with workers about them in a particular way, but this leaves open the question of whether workers can be encouraged to trust information about de-biased decision-making—unbiased algorithms in particular. On the other hand, my results suggest that beliefs about discrimination may be difficult to change. In that case, it will be important to study whether firms can mitigate the negative effects of perceived discrimination by encouraging workers to report such concerns and taking them seriously, as well as whether there are psychological or behavioral interventions that might help workers overcome the effects of perceived discrimination that later put them at a disadvantage.

References

Acikgoz, Yalcin, Kristl H. Davison, Maira Compagnone, and Matt Laske, "Justice Perceptions of Artificial Intelligence in Selection," *International Journal of Selection and Assessment*, 2020, *28* (4), 399–416.

Adams-Prassl, Abi, Kristiina Huttunen, Emily Nix, and Ning Zhang, "Violence Against Women at Work," *Quarterly Journal of Economics, Forthcoming.*

Agüero, Jorge M., Francisco Galarza, and Gustavo Yamada, "(Incorrect) Perceived Returns and Strategic Behavior among Talented Low-Income College Graduates," *AEA Papers and Proceedings*, 2023, *113*, 423–426.

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

Aksoy, Billur, Ian Chadd, and Boon Han Koh, "Sexual Identity, Gender, and Anticipated Discrimination in Prosocial Behavior," *European Economic Review*, 2023, *154*, 104427.

Alston, Mackenzie, "The (Perceived) Cost of Being Female: An Experimental Investigation of Strategic Responses to Discrimination," *Working Paper*, 2019.

Angeli, Deivis, Ieda Matavelli, and Fernando Secco, "Expected Discrimination and Job Search," *Working Paper*, 2024.

Anisman-Razin, Moran and Tamar Saguy, "Reactions to Tokenism: The Role of Individual Characteristics in Shaping Responses to Token Decisions," *European Journal of Social Psychology*, 2016, *46* (6), 716–731.

Avery, Mallory, Andreas Leibbrandt, and Joseph Vecci, "Does Artificial Intelligence Help or Hurt Gender Diversity? Evidence from Two Field Experiments on Recruitment in Tech," *Working Paper*, 2023.

Baker, Andrew, David F. Larcker, Charles McClure, Durgesh Saraph, and Edward M. Watts, "Diversity Washing," *Working Paper*, 2023.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 2016, *106* (6), 1437–1475.

Behaghel, Luc, Bruno Crépon, and Thomas Le Barbanchon, "Unintended Effects of Anonymous Résumés," *American Economic Journal: Applied Economics*, 2015, *7* (3), 1–27.

Benjamin, Daniel J., "Errors in Probabilistic Reasoning and Judgment Biases," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 2, Elsevier, 2019, pp. 69–186.

Bertrand, Marianne and Esther Duflo, "Field Experiments on Discrimination," in "Handbook of Economic Field Experiments," Vol. 1, Elsevier, 2017, pp. 309–393.

__ **and Sendhil Mullainathan**, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bigman, Yochanan, Kurt Gray, Adam Waytz, Mads Arnestad, and Desman Wilson, "Algorithmic Discrimination Causes Less Moral Outrage than Human Discrimination," *Journal of Experimental Psychology: General*, 2023, *152* (1), 4–27.

Blau, Francine D. and Lawrence M. Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg, "The Dynamics of Discrimination: Theory and Evidence," *American Economic Review*, 2019, *109* (10), 3395–3436.

__ , **Peter Hull, and Alex Imas**, "Systemic Discrimination: Theory and Measurement," *NBER Working Paper No. 29820*, 2022.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Beliefs About Gender," *American Economic Review*, 2019, *109* (3), 739–73.

__ , __ , __ , **Frederik Schwerter, and Andrei Shleifer**, "Memory and Representativeness," *Psychological Review*, 2021, *128* (1), 71–85.

Brecheisen, Jeremie, "Research: Where Employees Think Companies' DEIB Efforts Are Failing," *Harvard Business Review*, 2023.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Brown, Jason, Joe Burke, and Ashley Sauciuc, "Perceptions of Managerial Bias in Performance Evaluations: How Employee- Manager Demographics Influence Employee Bias Perceptions," *Working Paper*, 2024.

Campos-Mercade, Pol and Friederike Mengel, "Non-Bayesian Statistical Discrimination," *Management Science*, 2024, *70* (4), 2549–2567.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez, "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 2012, *102* (6), 2981–3003.

Chang, Edward H., Katherine L. Milkman, Dena M. Gromet, Robert W. Rebele, Cade Massey, Angela L. Duckworth, and Adam M. Grant, "The Mixed Effects of Online Diversity Training," *Proceedings of the National Academy of Sciences*, 2019, *116* (16), 7778–7783.

Charness, Gary, Ramón Cobo-Reyes, Simone Meraglia, and Ángela Sánchez, "Anticipated Discrimination, Choices, and Performance: Experimental Evidence," *European Economic Review*, 2020, *127*, 103473.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 2018, *21* (1), C1–C68.

Coate, Stephen and Glenn C. Loury, "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *The American Economic Review*, 1993, *83* (5), 1220–1240.

Coffman, Katherine B., Manuela R. Collis, and Leena Kulkarni, "Whether to Apply," *Harvard Business School Working Paper No. 20-062*, 2021.

Coffman, Katherine Baldiga, "Evidence on Self-Stereotyping and the Contribution of Ideas," *The Quarterly Journal of Economics*, 2014, *129* (4), 1625–1660.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider, "Social Comparison and Effort Provision: Evidence from a Field Experiment," *Journal of the European Economic Association*, 2014, *12* (4), 877–898.

Cowgill, Bo, "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening," *Working Paper*, 2020.

Dargnies, Marie-Pierre, Rustamdjan Hakimov, and Dorothea Kübler, "Aversion to Hiring Algorithms: Transparency, Gender Profiling, and Self-Confidence," *Working Paper*, 2022.

de Haan, Thomas, Theo Offerman, and Randolph Sloof, "Discrimination in the Labour Market: The Curse of Competition Between Workers," *The Economic Journal*, 2017, *127* (603), 1433–1466.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter De-

mand," *American Economic Review*, 2018, *108* (11), 3266–3302.

__ **, Lise Vesterlund, and Alistair J. Wilson**, "Experimenter Demand Effects," in "Handbook of Research Methods and Applications in Experimental Economics," Edward Elgar Publishing, 2019, pp. 384–400.

del Carmen Triana, María, Mevan Jayasinghe, and Jenna R. Pieper, "Perceived Workplace Racial Discrimination and its Correlates: A Meta-Analysis," *Journal of Organizational Behavior*, 2015, *36* (4), 491–513.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao, "Estimating Social Preferences and Gift Exchange at Work," *American Economic Review*, 2022, *112* (3), 1038–1074.

Dianat, Ahrash, Federico Echenique, and Leeat Yariv, "Statistical Discrimination and Affirmative Action in the Lab," *Working Paper*, 2020.

Doering, Laura, Jan Doering, and András Tilcsik, "'Was It Me or Was It Gender Discrimination?' How Women Respond to Ambiguous Incidents at Work," *Sociological Science*, 2023, *10*, 501–533.

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard, "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior," *American Economic Review*, 2019, *109* (2), 620–663.

__ **, Suresh Naidu, and Adam D. Reich**, "Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers," *NBER Working Paper No. 30441*, 2022.

Duquennois, Claire, "Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students," *American Economic Review*, 2022, *112* (3), 798–826.

Engel, Christoph, "Dictator Games: A Meta Study," *Experimental Economics*, November 2011, *14* (4), 583–610.

Ensher, Ellen A., Elisa J. Grant-Vallone, and Stewart I. Donaldson, "Effects of Perceived Discrimination on Job Satisfaction, Organizational Commitment, Organizational citizenship Behavior, and Grievances," *Human Resource Development Quarterly*, 2001, *12* (1), 53–72.

Fath, Sean, "When Blind Hiring Advances DEI — and When It Doesn't," *Harvard Business Review*, 2023.

Fehr, Ernst and Gary Charness, "Social Preferences: Fundamental Characteristics and Economic Consequences," *Working Paper*, 2023.

__ **, Lorenz Goette, and Christian Zehnder**, "A Behavioral Account of the Labor Market: The Role of Fairness Concerns," *Annual Review of Economics*, 2009, *1* (1), 355–384.

Fisher, R. A., *The Design of Experiments*, Oxford, England: Oliver & Boyd, 1935.

Flory, Jeffrey A., Andreas Leibbrandt, Christina Rott, and Olga Stoddard, "Increasing Workplace Diversity: Evidence from a Recruiting Experiment at a Fortune 500 Company," *Journal of Human Resources*, 2021, *56* (1), 73–92.

Folke, Olle and Johanna Rickne, "Sexual Harassment and Gender Inequality in the Labor Market," *The Quarterly Journal of Economics*, 2022, *137* (4), 2163–2212.

Fryer, Roland G., Jacob K. Goeree, and Charles A. Holt, "Experience-based Discrimination: Classroom Games," *The Journal of Economic Education*, 2005, *36* (2), 160–170.

Gagnon, Nickolas, Kristof Bosmans, and Arno Riedl, "The Effect of Gender Discrimination on Labor Supply," *Working Paper*, 2024.

Gallup Inc, "The Gig Economy and Alternative Work Arrangements," Technical Report 2018.

__ , "One in Four Black Workers Report Discrimination at Work," Technical Report 2021.

__ , "From Appreciation to Equity: How Recognition Reinforces DEI in the Workplace," 2023.

Glover, Dylan, Amanda Pallais, and William Pariente, "Discrimination as a Self-fulfilling Prophecy: Evidence from French Grocery Stores," *Quarterly Journal of Economics*, 2017, *132* (3), 1219–1260.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 2000, *90* (4), 715–741.

Goldsmith, Arthur H., Stanley Sedo, William Darity, and Darrick Hamilton, "The Labor Supply Consequences of Perceptions of Employer Discrimination During Search and On-the-job: Integrating Neoclassical Theory and Cognitive Dissonance," *Journal of Economic Psychology*, 2004, *25* (1), 15–39.

Gompers, Paul and Silpa Kovvali, "Diversity dividend," *Harvard Business Review*, 2018, pp. 72–77.

Hagmann, David, Gwendolin Sajons, and Catherine Tinsley, "Statistical Discrimination Against Minority Groups," *Working Paper*, 2022.

Hahn, Markus and Roger Wilkins, "Perceived Job Discrimination in Australia: Its Correlates and Consequences," *Australian Journal of Labour Economics*, 2013, *16* (1), 43–64.

Hrala, Josh, "Ways to Improve Diversity and Inclusion in the Workplace," 2022.

Jardina, Ashley, *White Identity Politics*, Cambridge University Press, 2019.

Jarrahi, Mohammad Hossein, Gemma Newlands, Min Kyung Lee, Christine T. Wolf, Eliscia Kinder, and Will Sutherland, "Algorithmic Management in a Work Context," *Big Data & Society*, 2021, *8* (2).

Jones, Kristen P., Chad I. Peddie, Veronica L. Gilrane, Eden B. King, and Alexis L. Gray, "Not So Subtle: A Meta-Analytic Investigation of the Correlates of Subtle and Overt Discrimination," *Journal of Management*, 2016, *42* (6), 1588–1613.

Kahneman, Daniel and Amos Tversky, "Subjective Probability: A Judgment of Representativeness," *Cognitive Psychology*, 1972, *3* (3), 430–454.

Kaibel, Chris, Irmela Koch-Bayram, Torsten Biemann, and Max Mühlenbock, "Applicant Perceptions of Hiring Algorithms - Uniqueness and Discrimination Experiences as Moderators," *Academy of Management Proceedings*, 2019.

Kardosh, Rasha, Asael Y. Sklar, Alon Goldstein, Yoni Pertzov, and Ran R. Hassin, "Minority Salience and the Overestimation of Individuals from Minority Groups in Perception and Memory," *Proceedings of the National Academy of Sciences of the United States of America*, 2022, *119* (12).

Kessler, Ronald C., Kristin D. Mickelson, and David R. Williams, "The Prevalence, Distribution, and Mental Health Correlates of Perceived Discrimination in the United States," *Journal of Health and Social Behavior*, 1999, *40* (3), 208–230.

King, Eden B., Michelle R. Hebl, Jennifer M. George, and Sharon F. Matusik, "Understanding Tokenism: Antecedents and Consequences of a Psychological Climate of Gender Inequity," *Journal of Management*, 2010, *36* (2), 482–510.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan, "Algorithmic Fairness," *AEA Papers and Proceedings*, 2018, *108*, 22–27.

__ , __ , __ , and Cass R. Sunstein, "Algorithms as Discrimination Detectors," *Proceedings of the National Academy of Sciences*, 2020, *117* (48), 30096–30100.

Kline, Patrick M., Evan K. Rose, and Christopher R. Walters, "Systemic Discrimination Among Large U.S.

Employers," *NBER Working Paper No. 29053*, 2021.

Lee, Min Kyung, "Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management," *Big Data & Society*, 2018, *5* (1).

Lepage, Louis-Pierre, Xiaomeng Li, and Basit Zafar, "Anticipated Gender Discrimination and Grade Disclosure," *NBER Working Paper No. 30765*, 2022.

Leslie, Lisa M., "Diversity Initiative Effectiveness: A Typological Theory of Unintended Consequences," *Academy of Management Review*, 2019, *44* (3), 538–563. Publisher: Academy of Management.

List, John A, Azeem M Shaikh, and Yang Xu, "Multiple Hypothesis Testing in Experimental Economics," *Experimental Economics*, 2019, *22* (4), 773–393.

Lundberg, Shelly J. and Richard Startz, "Private Discrimination and Social Intervention in Competitive Labor Market," *The American Economic Review*, 1983, *73* (3), 340–347.

Mukerjee, Swati, "Job Satisfaction in the United States: Are Blacks Still More Satisfied?," *The Review of Black Political Economy*, 2014, *41* (1), 61–81.

Mummolo, Jonathan and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," *American Political Science Review*, 2019, *113* (2), 517–529.

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

Newman, David T., Nathanael J. Fast, and Derek J. Harmon, "When Eliminating Bias isn't Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions," *Organizational Behavior and Human Decision Processes*, 2020, *160*, 149–167.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away from Competition? Do Men Compete Too Much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

__ , Carmit Segal, and Lise Vesterlund, "How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness," *Management Science*, 2012, *59* (1), 1–16.

Noble, Sean M., Lori L. Foster, and S. Bartholomew Craig, "The Procedural and Interpersonal Justice of Automated Application and Resume Screening," *International Journal of Selection and Assessment*, 2021, pp. 1–15.

Norton, Michael I. and Samuel R. Sommers, "Whites See Racism as a Zero-Sum Game That They Are Now Losing," *Perspectives on Psychological Science*, 2011, *6* (3), 215–218.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science*, 2019, *366*, 447–453.

Pager, Devah and David S. Pedulla, "Race, Self-Selection, and the Job Search Process," *American Journal of Sociology*, 2015, *120* (4), 1005–1054.

__ and Hana Shepherd, "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets," *Annual Review of Sociology*, 2008, *34*, 181–209.

Paluck, Elizabeth Levy, Roni Porat, Chelsey S. Clark, and Donald P. Green, "Prejudice Reduction: Progress and Challenges," *Annual Review of Psychology*, 2020, *72*.

Pascoe, Elizabeth A. and Laura Smart Richman, "Perceived Discrimination and Health: A Meta-analytic Review," *Psychological Bulletin*, 2009, *135* (4), 531–554.

Pew Research, "Gender Discrimination Comes in Many Forms for Today's Working Women," Technical Report 2017.

__ , "Racial and Ethnic Differences Stand Out in the U.S. Gig Workforce," Technical Report 2021.

__ , "The State of Gig Work in 2021," Technical Report 2021.

Porat, Roni and Elizabeth Levy Paluck, "Anger at Work," *Frontiers in Social Psychology*, 2024, *2*, 1337715.

Raghavan, Manish, Solon Barocas, Jon Kleinberg, and Karen Levy, "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 469–481.

Ridley, Matthew, "Mental Illness Discrimination," *Working Paper*, 2022.

Romano, Joseph P. and Michael Wolf, "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 2005, *73* (4), 1237–1282.

__ **and** __ , "Efficient Computation of Adjusted p-values for Resampling-based Stepdown Multiple Testing," *Statistics & Probability Letters*, 2016, *113*, 38–40.

Rosenbaum, Paul R., "Randomized Experiments," in Paul R. Rosenbaum, ed., *Observational Studies*, Springer Series in Statistics, New York, NY: Springer, 2002, pp. 19–66.

Ross, Lee, "The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process," in Leonard Berkowitz, ed., *Advances in Experimental Social Psychology*, Vol. 10, Academic Press, 1977, pp. 173–220.

__ **and Richard E. Nisbett**, *The Person and the Situation: Perspectives of Social Psychology* 1991.

Sanchez, Diana T., Kimberly E. Chaney, Sara K. Manuel, and Jessica D. Remedios, "Theory of Prejudice and American Identity Threat Transfer for Latino and Asian Americans," *Personality and Social Psychology Bulletin*, 2018, *44* (7), 972–983.

__ **,** __ **,** __ **, Leigh S. Wilton, and Jessica D. Remedios**, "Stigma by Prejudice Transfer: Racism Threatens White Women and Sexism Threatens Men of Color," *Psychological Science*, 2017, *28* (4), 445–461.

Schmader, Toni, Michael Johns, and Chad Forbes, "An Integrated Process Model of Stereotype Threat Effects on Performance," *Psychological Review*, 2008, *115* (2), 336–356.

Schmitt, Michael T., Nyla R. Branscombe, Tom Postmes, and Amber Garcia, "The Consequences of Perceived Discrimination for Psychological Well-being: A Meta-analytic Review," *Psychological Bulletin*, 2014, *140* (4), 921–948.

Small, Mario L. and Devah Pager, "Sociological Perspectives on Racial Discrimination," *Journal of Economic Perspectives*, 2020, *34* (2), 49–67.

Sockin, Jason, "Show Me the Amenity: Are Higher-Paying Firms Better All Around?," *Working Paper*, 2021.

Sorkin, Andrew Ross, Ravi Mattu, Bernhard Warner, Michael J. de la Merced, and Lauren Hirsch, "Why Corporate America Is Worried About Affirmative Action," *The New York Times*, June 2023.

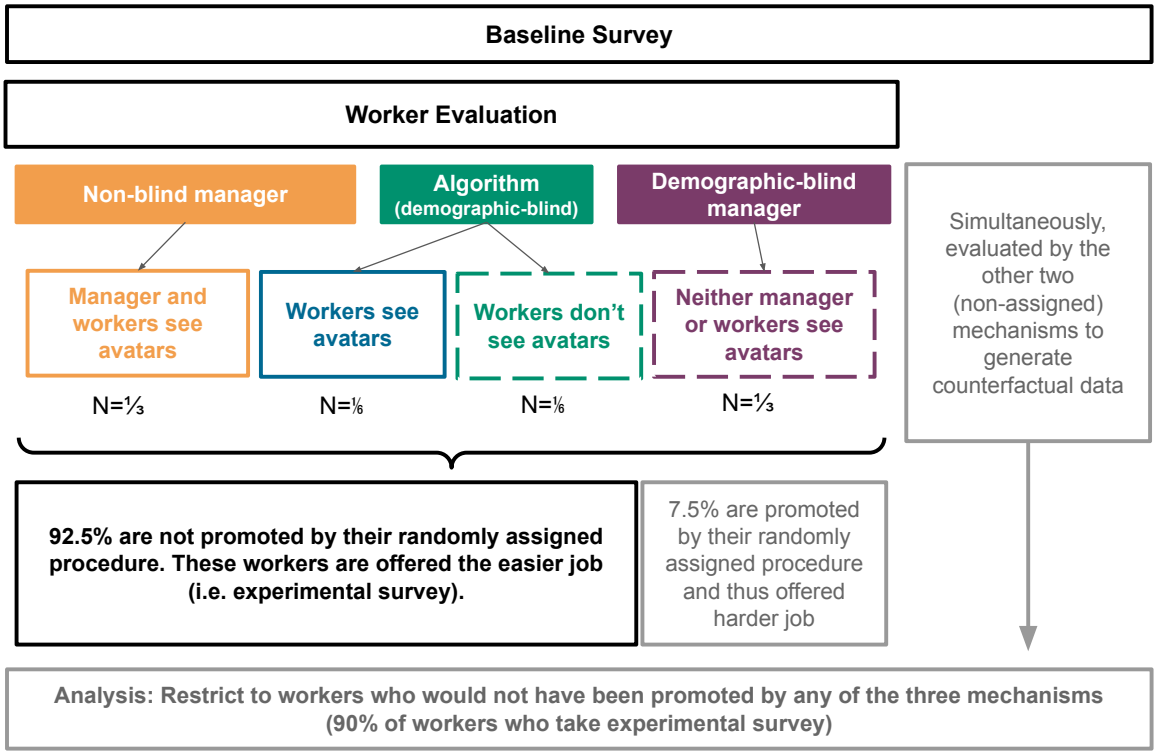
Stainback, Kevin and Matthew Irvin, "Workplace Racial Composition, Perceived Discrimination, and Organizational Attachment," *Social Science Research*, 2012, *41* (3), 657–670.

Thoits, Peggy A., "Stress and Health: Major Findings and Policy Implications," *Journal of Health and Social Behavior*, 2010.

Vick, S. Brooke, Mark D. Seery, Jim Blascovich, and Max Weisbuch, "The Effect of Gender Stereotype Activation on Challenge and Threat Motivational States," *Journal of Experimental Social Psychology*, 2008, *44* (3), 624–630.

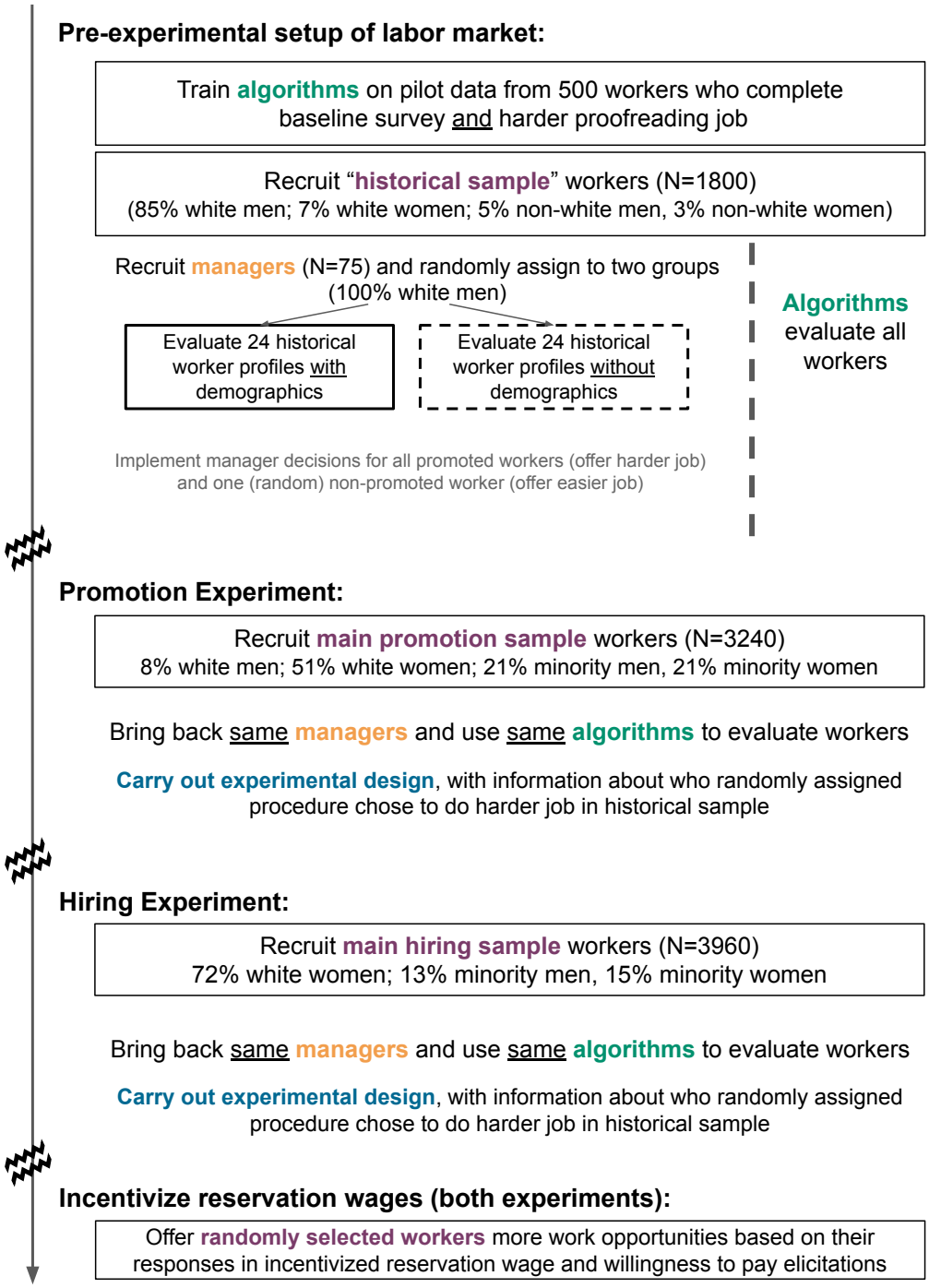
Zhang, Lixuan and Christopher Yencha, "Examining Perceptions Towards Hiring Algorithms," *Technology in Society*, 2022, *68*, 101848.

Figure 1: Experimental design of the promotion experiment



Note: This figure describes the design of the promotion experiment. The baseline survey includes grammar, science, and spelling quizzes and confidence in each, demographics, work history, and past experiences of and beliefs about the prevalence of discrimination. The set-up for workers in the historical sample was similar, except that only one worker in the non-promoted group was offered the easier job. The design of the hiring experiment was the same except that workers were assigned to be evaluated by one of four procedures, and 97.5% of workers were not hired under their randomly assigned procedure (see Figure 5).
[Click to return to Section 2 (Research Design)]

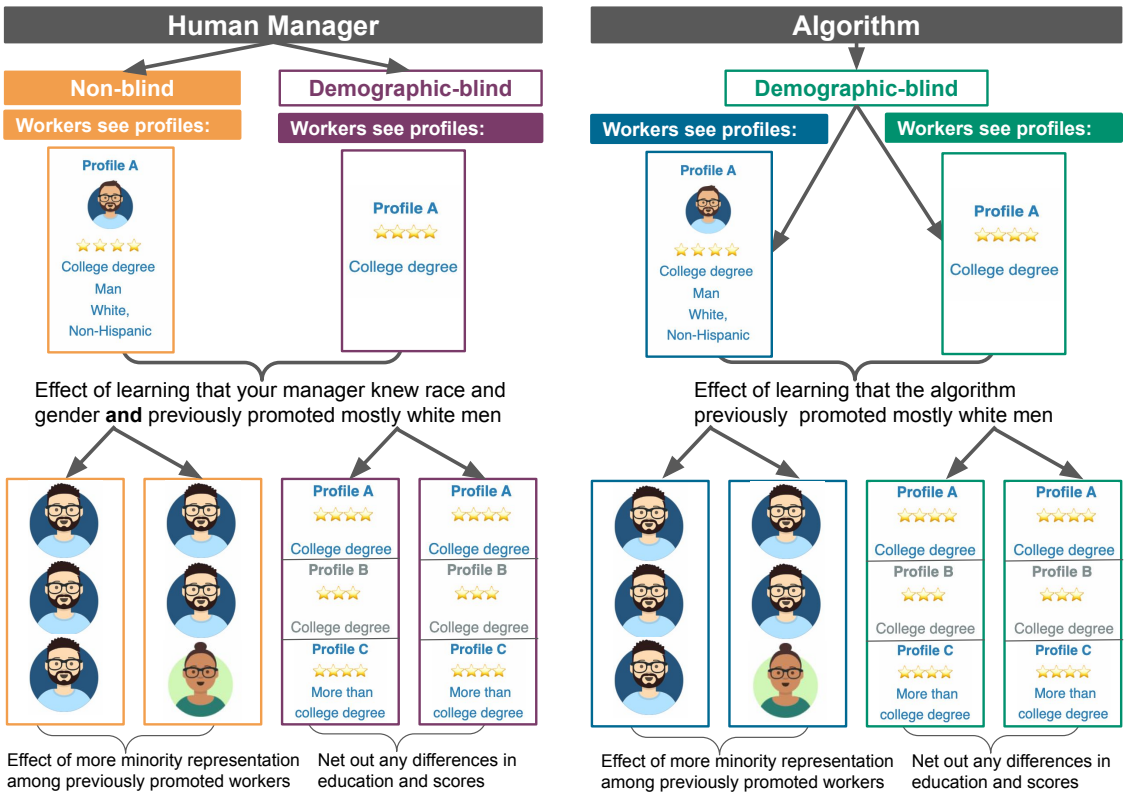
Figure 2: Timeline of study



Note: This figure describes the dynamics of the labor market created on Prolific to generate repeated interactions with workers and managers. The pre-experimental setup describes the recruitment of the historical sample and managers. The same managers were brought back to evaluate the workers in the experimental samples. The main difference between the historical and experimental samples was the demographic composition, in order to approximate a scenario in which under-represented workers saw that their manager had mostly promoted workers from the majority group in the past. After the conclusion of the experiment, randomly-selected workers are offered additional proofreading jobs that implement the scenarios in the reservation wage and willingness to pay elicitations.

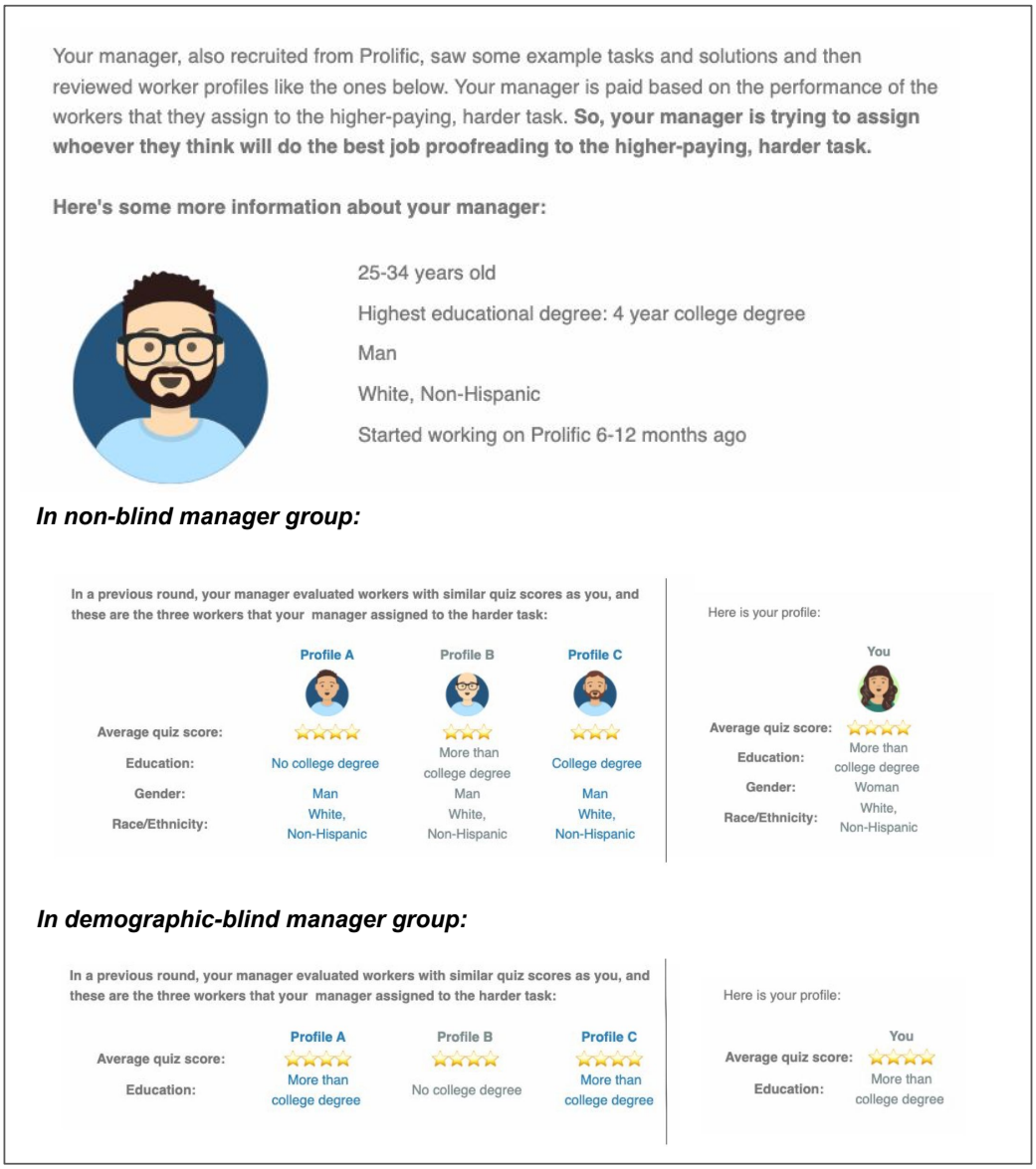
[Click to return to Section 2 (Research Design)]

Figure 3: Treatment arms in the promotion experiment



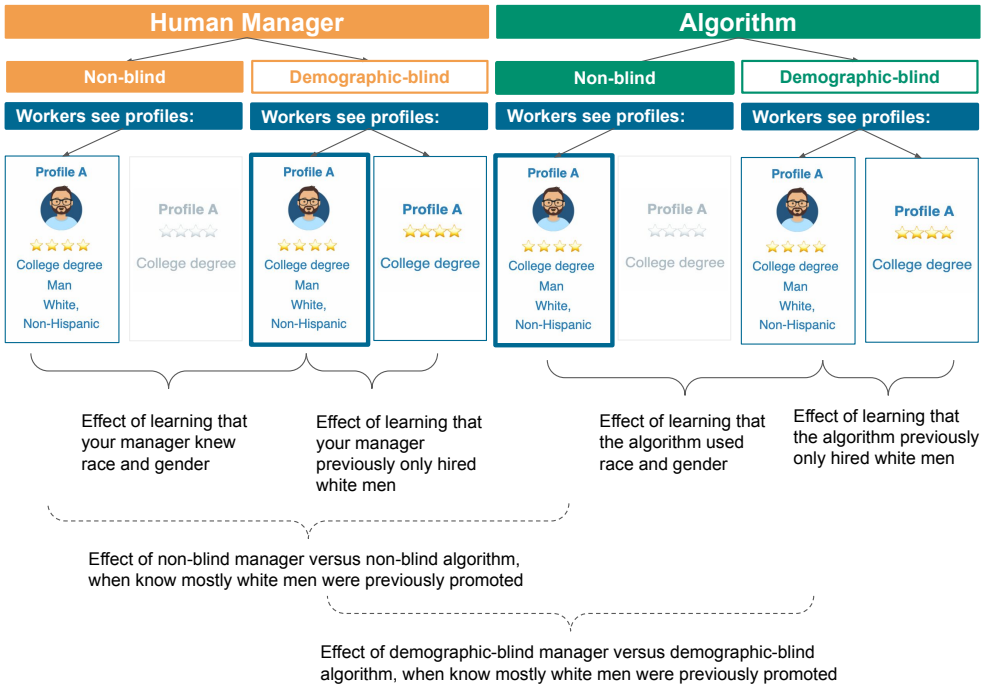
Note: This figure illustrates the treatment arms in the promotion experiment and how they are used to answer various research questions. Randomization is indicated by gray arrows. In the manager sample, workers are randomly assigned to managers. Some previously promoted three white men, some promoted two white men and someone else; this is uncorrelated with other manager characteristics. In the algorithm sample, workers are randomly assigned to groups of workers jointly evaluated by the algorithm (analogously to how the managers jointly evaluated groups of workers). Similarly, in some groups, three white men were previously promoted, and in others, it was two white men and someone else.
[Click to return to Section 2 (Research Design)]

Figure 4: Treatment variation in manager groups



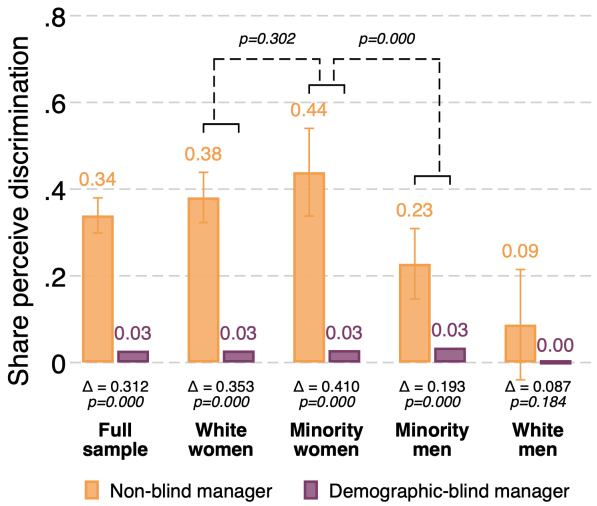
Note: This figure shows what workers saw about their manager, regardless of whether their manager saw avatars, race, and gender or not, and the information they saw about the workers their manager previously promoted in the non-blind manager group and demographic-blind manager group. Figure A1 shows the analogous implementation of the algorithm arms.
[Click to return to Section 2 (Research Design)]

Figure 5: Treatment arms of the hiring experiment



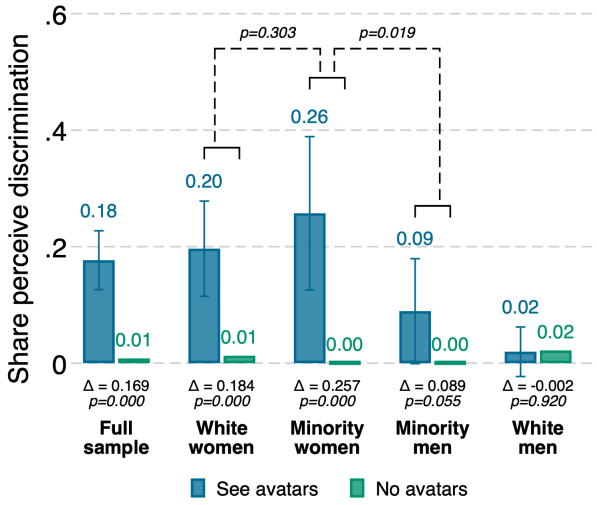
Note: This figure illustrates the treatment arms in the hiring experiment, which replicates the four original treatment arms and adds two new arms, bolded. The grayed-out boxes indicate arms that would be included in the full 2x2x2 factorial design, but are not truthfully or realistically implementable in a way parallel to the other arms. Workers are randomized equally across the six arms, as indicated by the gray arrows. The figure shows how comparisons between the arms are used to answer various research questions.
[Click to return to Section 2 (Research Design)]

Figure 6: Perceived discrimination in the manager arms of the promotion experiment



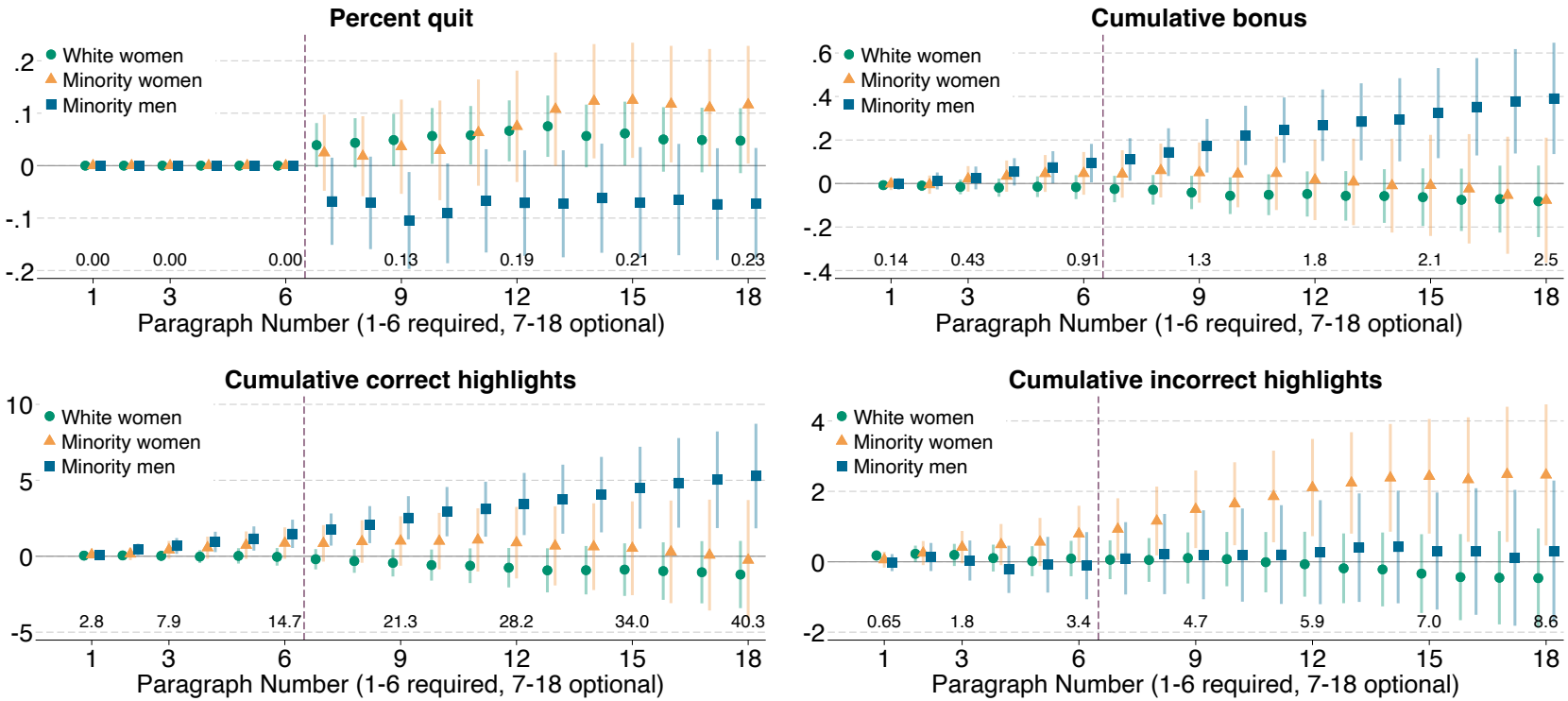
Note: This figure plots the share of workers in the experimental sample of the promotion experiment who perceive discrimination using the main measure (whether a worker mentions demographics in their response to the open-ended question, *"what needed to be different about your profile in order to be assigned to the harder task?"*) first for all workers and then separately by racial minority status × gender in the two manager arms. The plotted means are calculated from the estimated coefficients in Equation 1 separately for each subgroup, controlling for quiz scores, education, income, age, marital and parental status, race, gender, quiz-score group fixed effects, and the educational and previous-performance composition of the previously-promoted workers each worker saw. The figure restricts to workers who would not have been promoted under any promotion procedure. For breakdowns by race × gender, see Appendix Figure A4. 95 percent confidence intervals calculated with standard errors robust to heteroskedasticity are indicated by the black bars.
[Click to return to Section 3.2 (Effects of Perceived Discrimination)]

Figure 7: Perceived discrimination in the algorithm arm of the promotion experiment



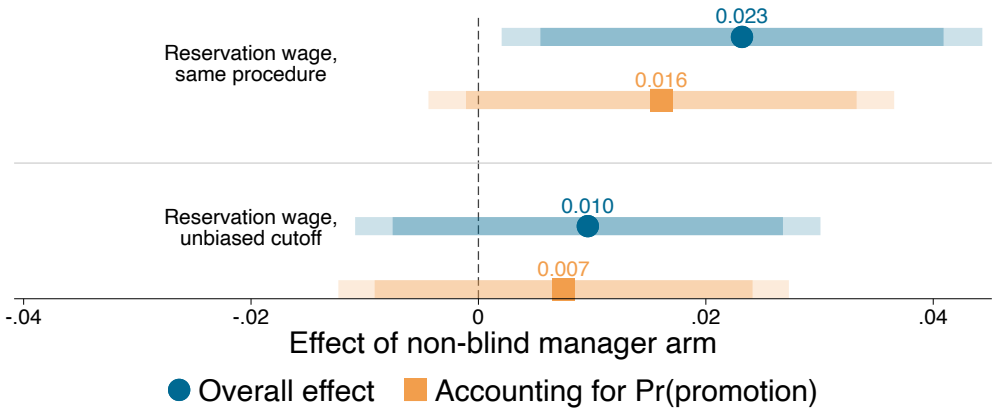
Note: This figure is analogous to Figure 6 but compares perceived discrimination rates in the two sub-arms of the algorithm treatment arm. The sample is workers in the algorithm arm who would not have been promoted under any of the three procedures.
[Click to return to Section 3.2 (Effects of Perceived Discrimination)]

Figure 8: Treatment effects on retention and performance by gender × race



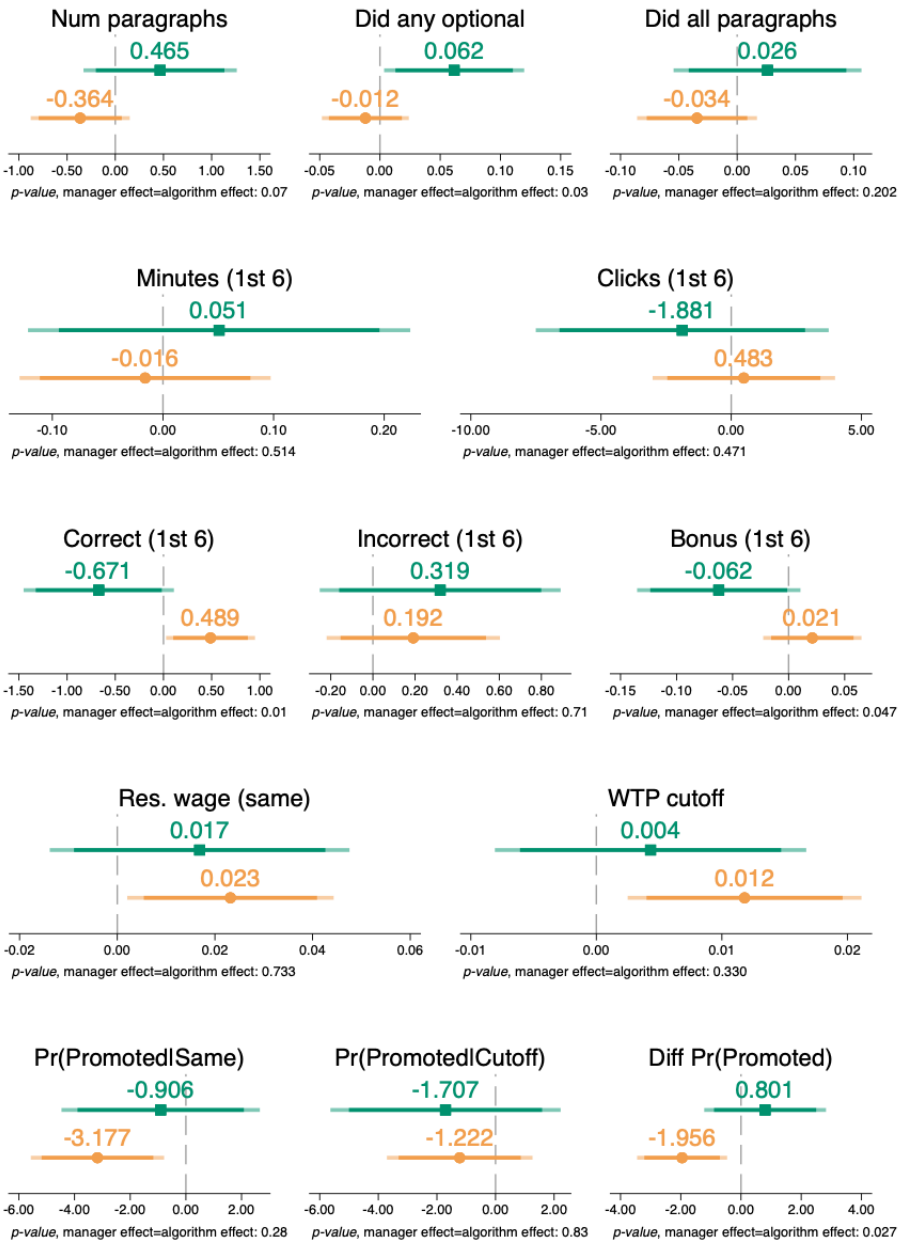
Note: This figure plots the effect of being in the non-blind manager arm of the promotion experiment on retention and performance measures, each cumulatively by the given paragraph. Equation 1 is estimated interacting the treatment indicator with indicators for each race/gender group; controls are the same as in Table 2. 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

Figure 9: Decomposing the effects on reservation wages in the promotion experiment



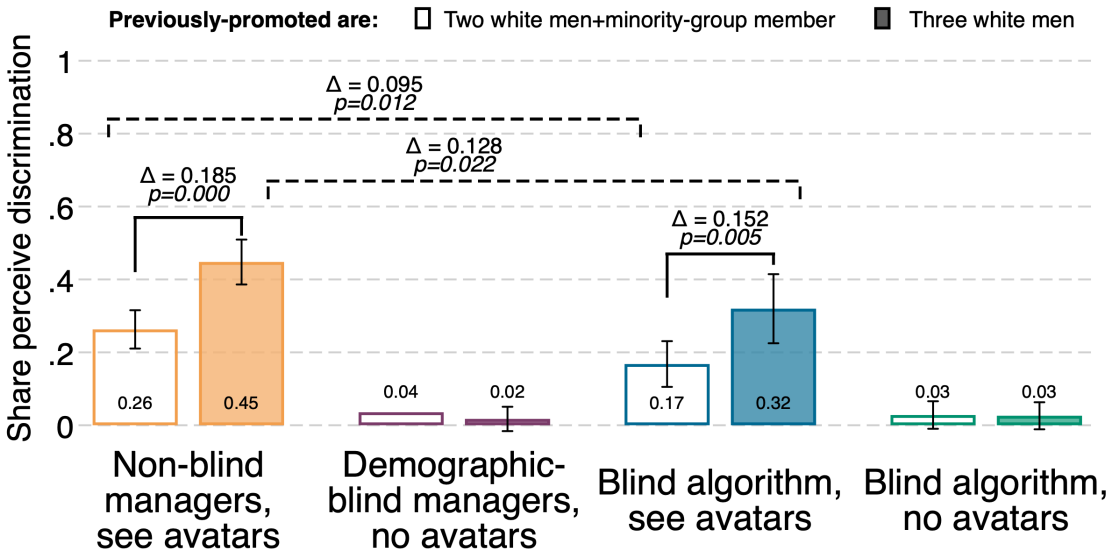
Note: This figure plots the effect of being in the non-blind manager arm relative to the demographic-blind manager arm of the promotion experiment on workers' reservation wages when they will be evaluated by the same manager as in the experiment (top panel) or by an unbiased cutoff rule in baseline quiz scores (bottom panel). Effects are plotted first on the overall measure of reservation wages, then on the measure of reservation wages that accounts for effects on beliefs about the likelihood of promotion, isolating the extent to which the effect is due to anticipated disutility. The sample and controls are the same as in Table 4. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.4.2 (Effects on Future Labor Supply, Mechanisms)]

Figure 10: Effects of perceived algorithmic discrimination on worker behavior



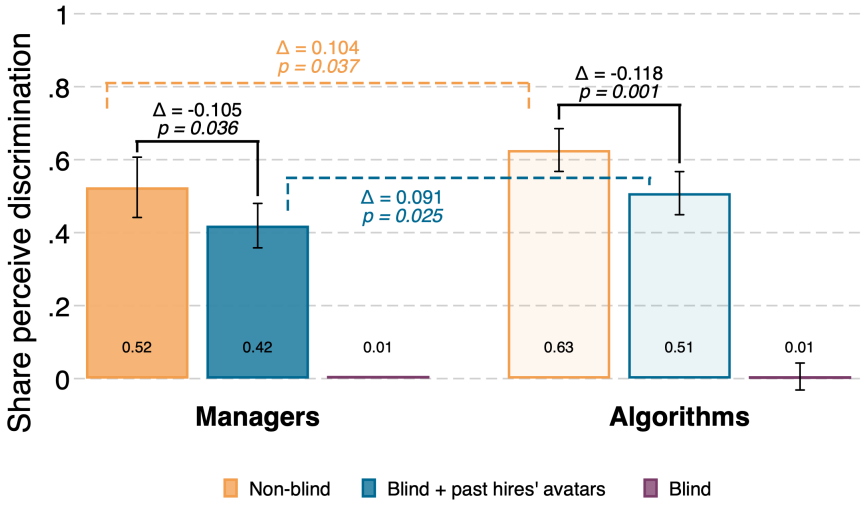
Note: This figure plots the treatment effects of learning that previously-promoted workers were mostly white men in the algorithm arm of the promotion experiment, and compares this effect to the effect of learning that one's manager knew demographic information *and* previously promoted mostly white men. The outcomes and control variables are the same as in Tables 2, 3, 4, and 6. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.5 (Effects of Perceived Algorithmic Discrimination)]

Figure 11: Effects of seeing one previously-promoted minority-group worker



Note: This figure plots the share in each treatment arm of the promotion experiment perceiving discrimination by the main measure, separately by whether workers (would) see three white men previously promoted or not. Those who do not see three white men see two white men and one member of a minority group. Workers in the demographic-blind manager arm and those in the algorithm arm who do not see worker avatars are split by whether they *would* be shown three white men or not. The controls are the same as Figure 6 and the sample is workers in the promotion experiment who would not have been promoted under any procedure and are not white men. *p*-values and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 4 (Effects of Minority-Group Representation)]

Figure 12: Perceived discrimination in the hiring experiment



Note: This figure plots the share of workers perceiving discrimination in each treatment arm of the hiring experiment, using the main measure of perceived discrimination. The sample is restricted to workers who would not have been hired under any hiring procedure and workers who see three previous hires who are all white men (a random subsample within each treatment arm). Shares are calculated via regressions with the same controls as Figure 6. *p*-values and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 5 (Effects of Hiring Policies)]

Table 1: Summary statistics and comparison to ACS sample

	Experimental Samples		ACS 2021, ages 18-71	
	Promotion	Hiring	All	Subsample
	(1)	(2)	(3)	(4)
Race and Gender:				
Asian men	0.08	0.03	0.03	0.05
Black men	0.07	0.05	0.06	0.08
Hispanic men	0.07	0.05	0.09	0.15
White men	0.08	0.00	0.30	0.00
Other race men	–	–	0.02	0.04
Asian women	0.05	0.03	0.03	0.05
Black women	0.07	0.07	0.06	0.09
Hispanic women	0.07	0.06	0.09	0.12
White women	0.51	0.72	0.29	0.40
Other race women	–	–	0.02	0.03
Married	0.54	0.58	0.50	0.48
Kids	0.43	0.46	0.39	0.44
Education:				
Less than high school	0.01	0.00	0.10	0.09
High school graduate	0.13	0.09	0.27	0.24
Some college but no degree	0.27	0.25	0.21	0.21
2 year college degree	0.09	0.09	0.09	0.09
4 year college degree	0.37	0.40	0.21	0.23
Professional or Masters degree	0.11	0.15	0.11	0.13
Doctorate	0.01	0.02	0.01	0.02
Income:				
Less than $20,000	0.14	0.11	0.12	0.09
$20,000-$40,000	0.22	0.17	0.15	0.15
$40,000-$60,000	0.18	0.19	0.14	0.14
$60,000-$80,000	0.15	0.16	0.12	0.13
$80,000-$100,000	0.10	0.13	0.10	0.11
$100,000-$120,000	0.07	0.07	0.08	0.09
$120,000-$140,000	0.04	0.06	0.06	0.07
$140,000-$160,000	0.04	0.04	0.05	0.05
More than $160,000	0.06	0.07	0.18	0.18
Age:				
18-24	0.17	0.15	0.13	0.13
25-34	0.32	0.33	0.20	0.24
35-44	0.22	0.22	0.19	0.23
45-54	0.14	0.16	0.18	0.20
55-64	0.11	0.10	0.19	0.16
65 or older	0.05	0.04	0.11	0.04
Employment:				
Currently employed outside Prolific	0.67	0.69	–	–
Ever employed outside Prolific	0.96	0.97	–	–
In the labor force	–	–	0.72	1.00
Employed	–	–	0.67	0.93
Employed if in labor force	–	–	0.94	0.93
N	2080	1934	2216940	1005688

Note: The samples in columns 1 and 2 are the experimental sample workers who would not have been selected by any of the 3 or 4 procedures in the promotion and hiring experiments, respectively, each combining all treatment arms. The sample in column 3 is respondents age 18-71 in the 2021 ACS (the minimum and 99th percentile age in experimental sample, respectively). Column 4 further restricts to women and racial-minority men currently in the labor force. All variables are indicators, and means are presented for each variable. ACS estimates weight by the *perwt* variable from IPUMS; experimental sample estimates are unweighted.
[Click to return to Section 2 (Research Design)]

Table 2: Effects on retention

	Num paragraphs	Did any optional	Did all paragraphs
	(1)	(2)	(3)
Non-blind manager,	-0.364	-0.012	-0.034
see avatars	(0.262)	(0.018)	(0.026)
N	1293	1293	1293
Control mean	15.857	0.911	0.771

Note: This table reports treatment effects of being in the non-blind manager arm in the promotion experiment on retention (Equation 1. The first outcome is the number of paragraphs completed; workers had to do at least six to receive their participation payment and were given the option to quit after each subsequent paragraph. They could proofread up to eighteen paragraphs. The second and third outcomes are indicators for doing more than six paragraphs and finishing all eighteen, respectively. The sample is those evaluated by a manager and the presented coefficient is on the indicator for being in the non-blind manager arm rather than the demographic-blind manager arm, restricting to workers who would not have been promoted under any evaluation procedure in the promotion experiment and workers who are not white men. All regressions control for quiz scores, education, income, age, marital and parental status, race, gender, quiz-score group fixed effects, and the educational and previous-performance composition of the previously-promoted workers each worker saw. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

Table 3: Effects on effort and performance in the first six (required) paragraphs

	Minutes	Clicks	Correct	Incorrect	Bonus
	(1)	(2)	(3)	(4)	(5)
Non-blind manager,	-0.016	0.483	0.489**	0.192	0.021
see avatars	(0.058)	(1.786)	(0.235)	(0.210)	(0.022)
N	1295	1295	1295	1295	1295
Control mean	4.946	35.981	14.705	3.412	0.910

Note: This table reports treatment effects of being in the non-blind manager arm in the promotion experiment on effort and performance. The effort outcomes in columns 1 and 2 are the minutes spent and number of clicks on the first six required paragraphs. The performance outcomes in columns 3, 4, and 5 are the number of correct highlights, number of incorrect highlights, and bonus earned in dollars in the first six paragraphs. Samples and specifications are the same as Table 2. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

Table 4: Effects on future labor supply

	Res. wage (same procedure)	WTP unbiased procedure
	(1)	(2)
Non-blind manager,	0.023**	0.012**
see avatars	(0.011)	(0.005)
N	1247	1235
Control mean	0.265	-0.015

Note: This table reports treatment effects of being in the non-blind manager arm in the promotion experiment on measures of future labor supply. The first outcome is the reservation wage (per well-proofread paragraph) to be evaluated again under the same procedure as in the experiment. The second is the reservation wage from column 1 minus the reservation wage when an unbiased cutoff rule is used to decide future promotions. Samples and specifications are the same as Table 2, with additional workers dropped who did not respond to the MPLs in sensical ways (see Appendix B). Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.4 (Effects on Future Labor Supply)]

Table 5: Effects on sharing with and avoidance of manager

	Cooperative task RW	WTP new manager	Shared zero	Amt bonus shared
	(1)	(2)	(3)	(4)
Non-blind manager, see avatars	0.007	0.050**	0.049	-0.285
	(0.015)	(0.021)	(0.030)	(0.244)
N	1256	957	1289	1289
Control mean	0.339	0.218	0.531	2.813

Note: This table reports treatment effects of being in the non-blind manager arm in the promotion experiment on measures of cooperation with and generosity towards managers. The first and second outcomes are workers' reservation wage per paragraph to do a future collaborative task with their manager and their willingness to pay (as a reduction in a $1 wage per paragraph) to be able to choose their own manager for the collaborative task. The third outcome is an indicator for sharing none of the 20-dollar surprise bonus with their manager and the fourth is the amount of the bonus that they share in dollars. Samples and specifications are the same as in Table 2, with additional workers dropped who did not respond to the MPLs in sensical ways (see Appendix B). Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.4 (Effects on Future Labor Supply)]

Table 6: Effects on beliefs about future promotion

	Pr(Promoted|Same)	Pr(Promoted|Cutoff)	Diff Pr(Promoted)
	(1)	(2)	(3)
Non-blind manager, see avatars	-3.177***	-1.222	-1.956**
	(1.220)	(1.267)	(0.762)
N	1291	1291	1291
Control mean	47.397	48.411	-1.014

Note: This table reports treatment effects of being in the non-blind manager arm in the promotion experiment on beliefs about the likelihood of promotion in the future. In the first two columns, the outcomes are the believed probability of being promoted in a future round of the study where evaluation is under the same procedure as in the experiment and using a cutoff rule in baseline quiz scores, respectively. The third outcome is the difference between the two. Samples and specifications are the same as Table 2. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.4.2 (Effects on Future Labor Supply, Mechanisms)]

Online Appendix:

Causes and Consequences of Perceived Workplace Discrimination

Hannah Ruebeck

Table of Contents

A Supplementary figures and tables

B Robustness analyses

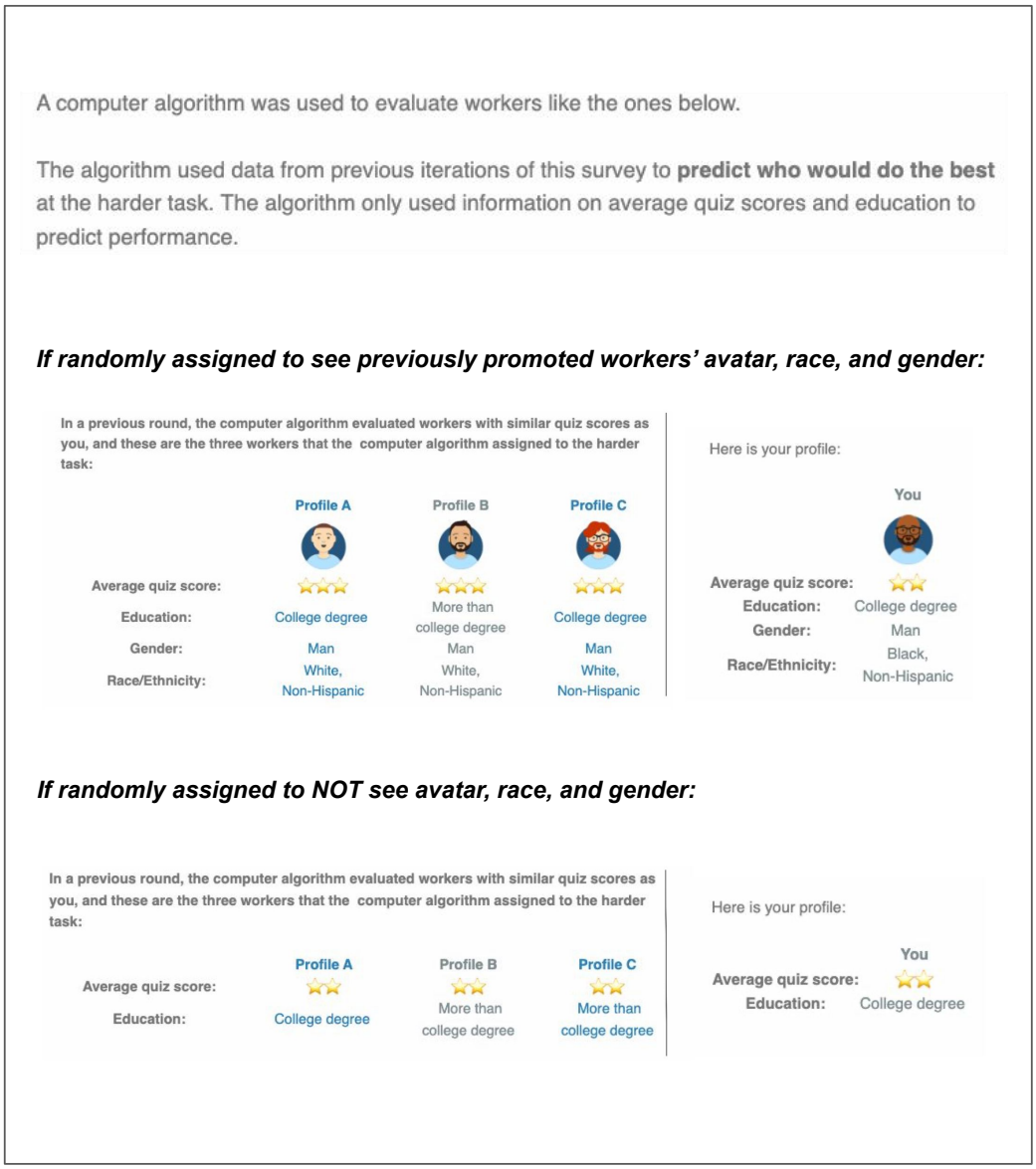
Online Supplement (LINK)

Contents:

Promotion procedure details
Survey implementation
Coding the text data
Variable definitions
Stereotype survey design and results
IV specifications
Effects on future labor supply in the hiring experiment (replication)

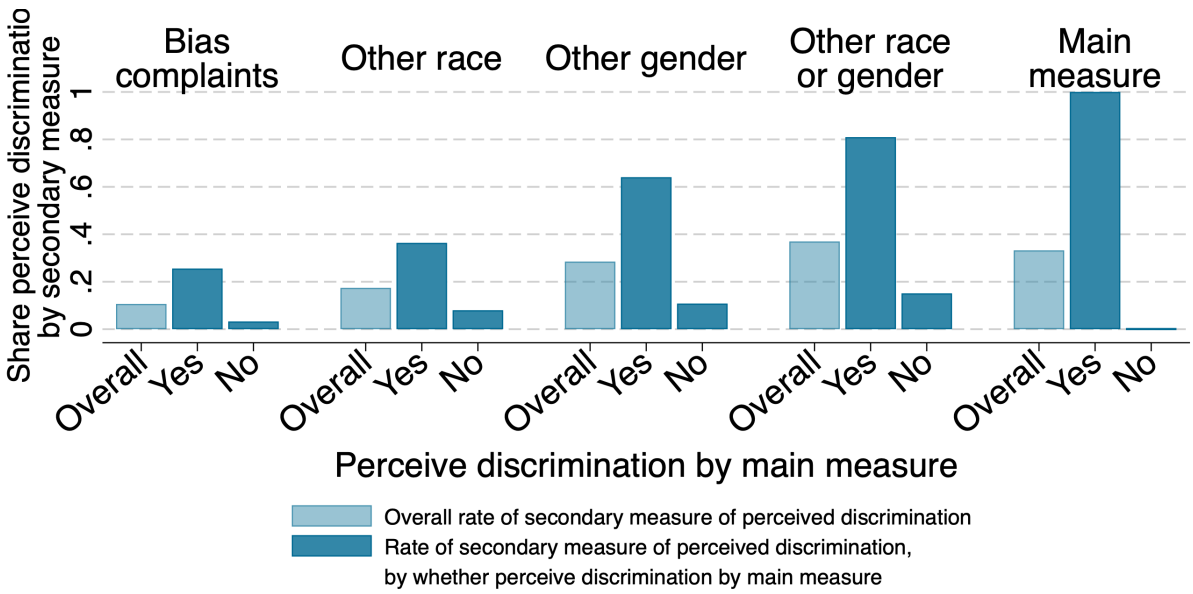
A Supplementary figures and tables

Figure A1: Treatment variation in algorithm groups



Note: This figure shows what workers saw about the algorithm and the information they saw about the workers the algorithm previously promoted depending on whether they were randomly assigned to see those workers' avatars, race, and gender or not. This can also be compared to Figure 4 which shows the analogous implementation of the manager arms.
[Click to return to Section 2 (Research Design)]

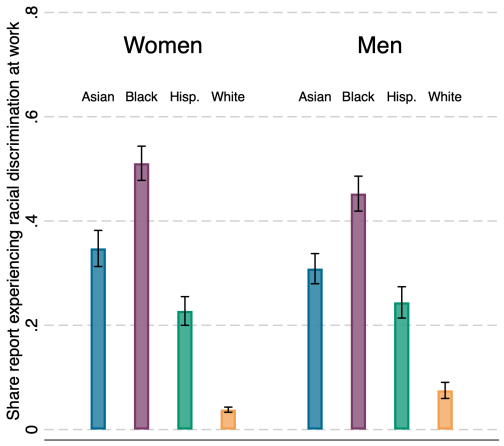
Figure A2: Correlation of perceived discrimination measures



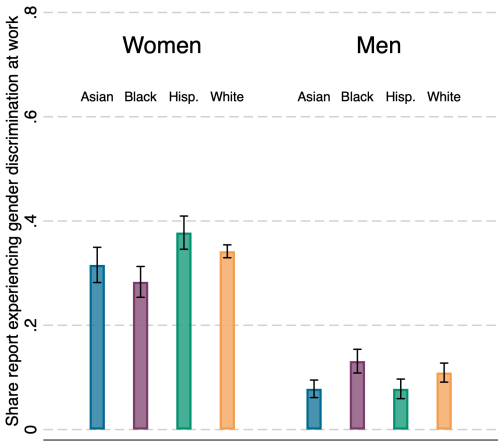
Note: This figure shows the correlation between the primary and secondary measures of perceived discrimination and the overall frequency of each in the non-blind manager arms of the promotion and hiring experiments. The primary measure is an indicator for whether a worker mentions demographics in their response to the open-ended question, *"what needed to be different about your profile in order to be assigned to the harder task?"* The secondary measures are an indicator for mentioning discrimination or bias in a open-ended response describing their complaint about the promotion procedure if they said they had one, an indicator for saying they thought they would have been promoted if they had a different race, the same for gender, and an indicator that is the maximum of the two. The light blue bars plot the share of workers who perceive discrimination by each measure in the full sample, and then the dark blue bars split the sample into workers who perceive discrimination by the main measure and those who don't.
[Click to return to Section 2 (Research Design)]

Figure A3: Perceived discrimination at work in the past (baseline survey)

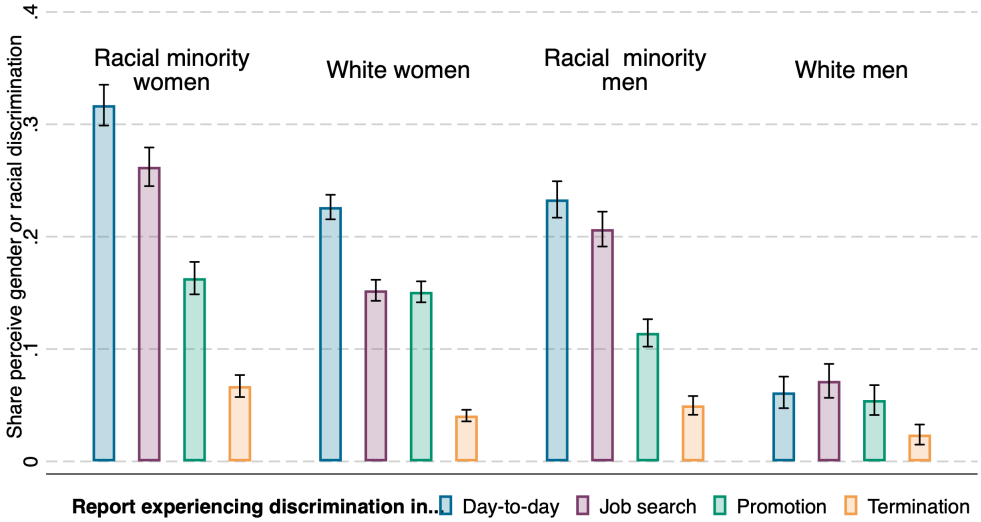
(a) Racial discrimination



(b) Gender discrimination



(c) By work setting

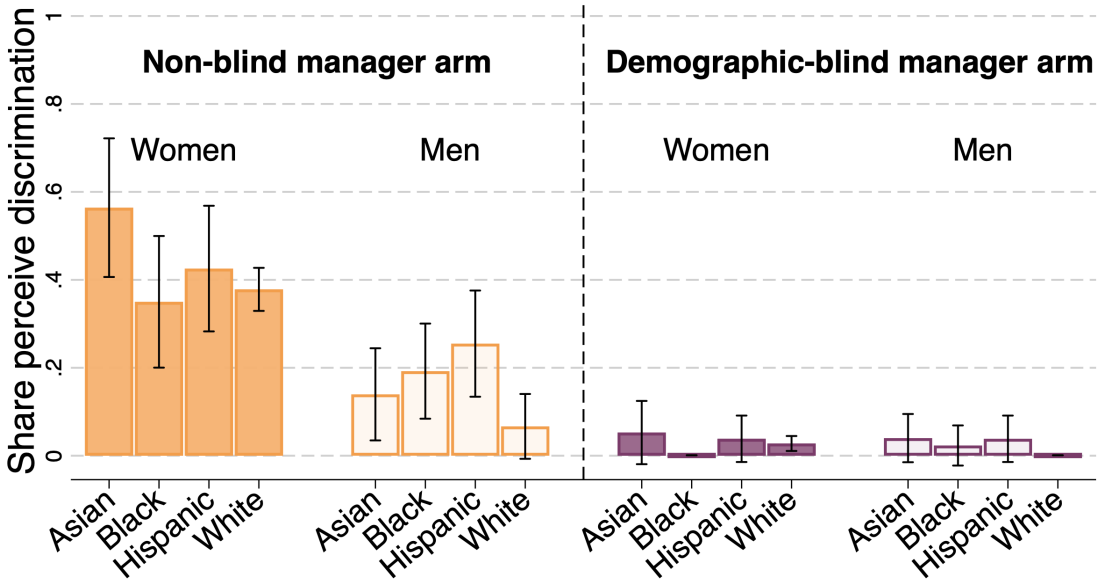


Note: The sample is 3,102 workers recruited for the promotion experiment who have ever held a job outside Prolific (96% of recruited workers). The share of each group who reported having experienced discrimination at work in the past is plotted with plus or minus the standard error in this sample indicated by black bars. Panels A and B pool discrimination during job search, promotion, termination, and daily work activities and separate racial and gender discrimination. Panel C pools racial and gender discrimination and separates the work setting. The share of each group who reported having experienced discrimination at work in the past is plotted with plus or minus the standard error in this sample indicated by black bars, separately for whether workers report perceived discrimination during daily work activities, job search, promotion, and termination.

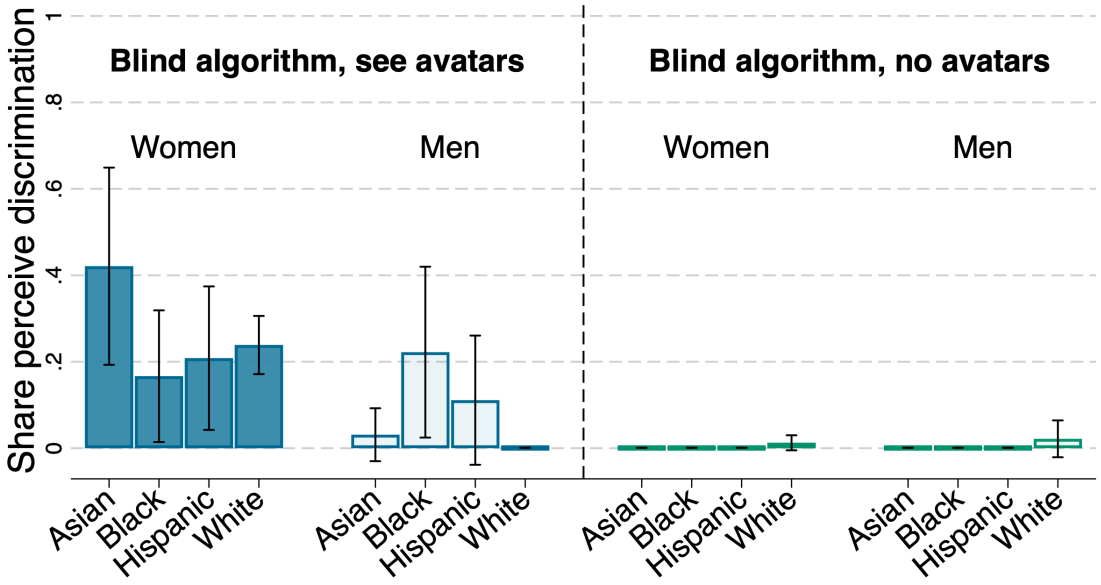
[Click to return to Section 1 (Introduction)]
[Click to return to Section 2.5 (Generalizability)]

Figure A4: Perceived discrimination by race × gender

Panel A: Manager sample



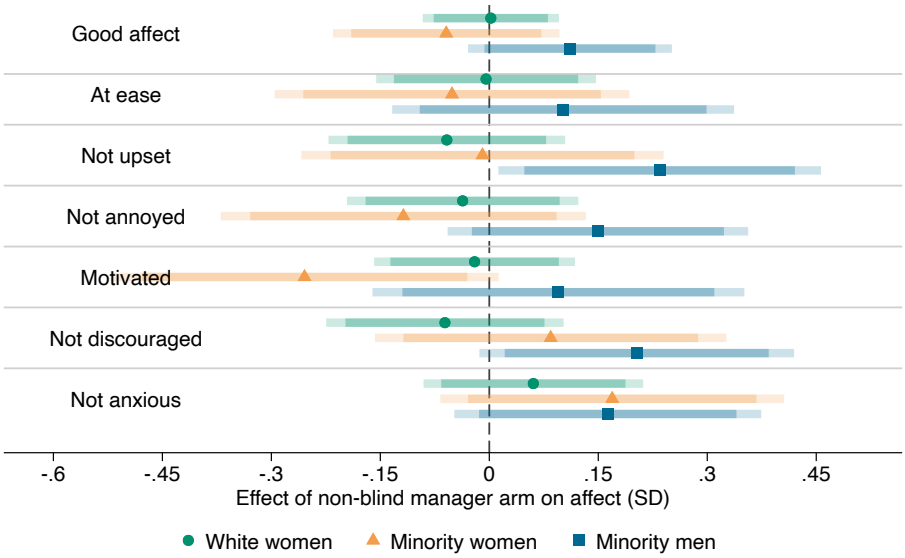
Panel B: Algorithm sample



Note: This figure plots the share of workers in the experimental sample who perceived discrimination using the main measure (mentioning demographics in their response to the open-ended question, "*what needed to be different about your profile to be assigned to the harder, higher-paying task?"*) separately by race × gender in the promotion experiment, separately by treatment arm. The sample restricts to workers who would not have been promoted under any procedure. 95 percent confidence intervals are indicated by the black bars.

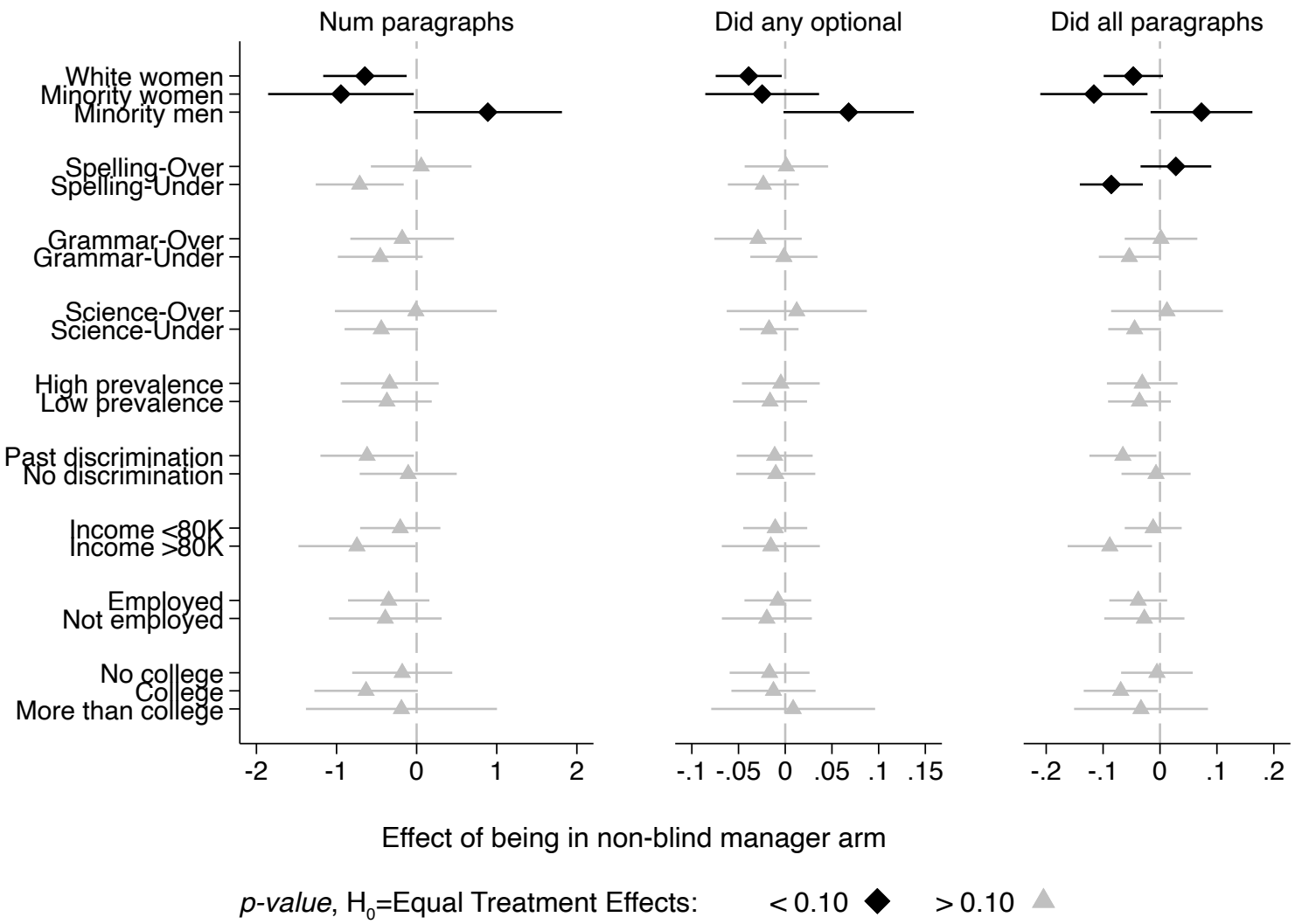
[Click to return to Section 3.2 (Effects of Perceived Discrimination)]

Figure A5: Effects on psychological well-being, gender and race heterogeneity



Note: This figure plots the effect of being in the non-blind manager arm relative to the demographic-blind manager arm in the promotion experiment on emotional states, measured with an affective well-being scale and reported in standard deviations, separately by race and gender. The regressions restrict to workers who would not have been promoted under any promotion procedure who are not white men. Controls are the same as in Table 2. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

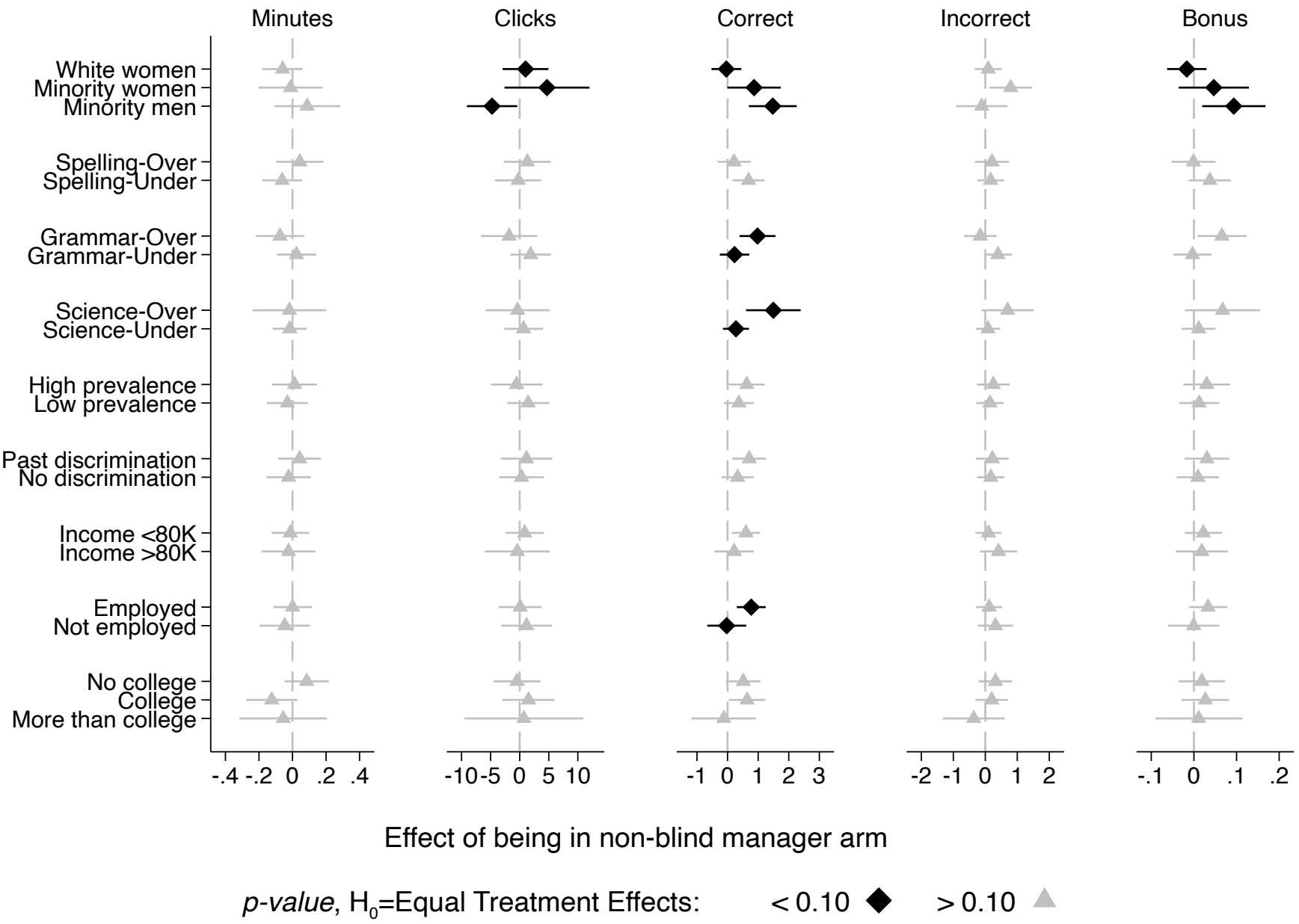
Figure A6: Heterogeneity by all potential mechanism variables, retention outcomes



Note: This figure estimates the effect of treatment separately by race, over/under-confidence, beliefs about the prevalence of discrimination and experiencing discrimination in the past, income, employment, and education. Each pair or triple of coefficients is estimated in one regression by interacting treatment with each mutually-exclusive indicator. The indicator for high beliefs about the prevalence of discrimination is 1 if a worker has above-median beliefs about the likelihood of discrimination against their own race *times* gender group compared to others in their race and gender group. Controls are the same as in the main specification, with the main effects of the heterogeneity variables added as controls. **90** percent confidence intervals are shown, calculated with standard errors are robust to heteroskedasticity. The outcomes are the same as in Table 2.

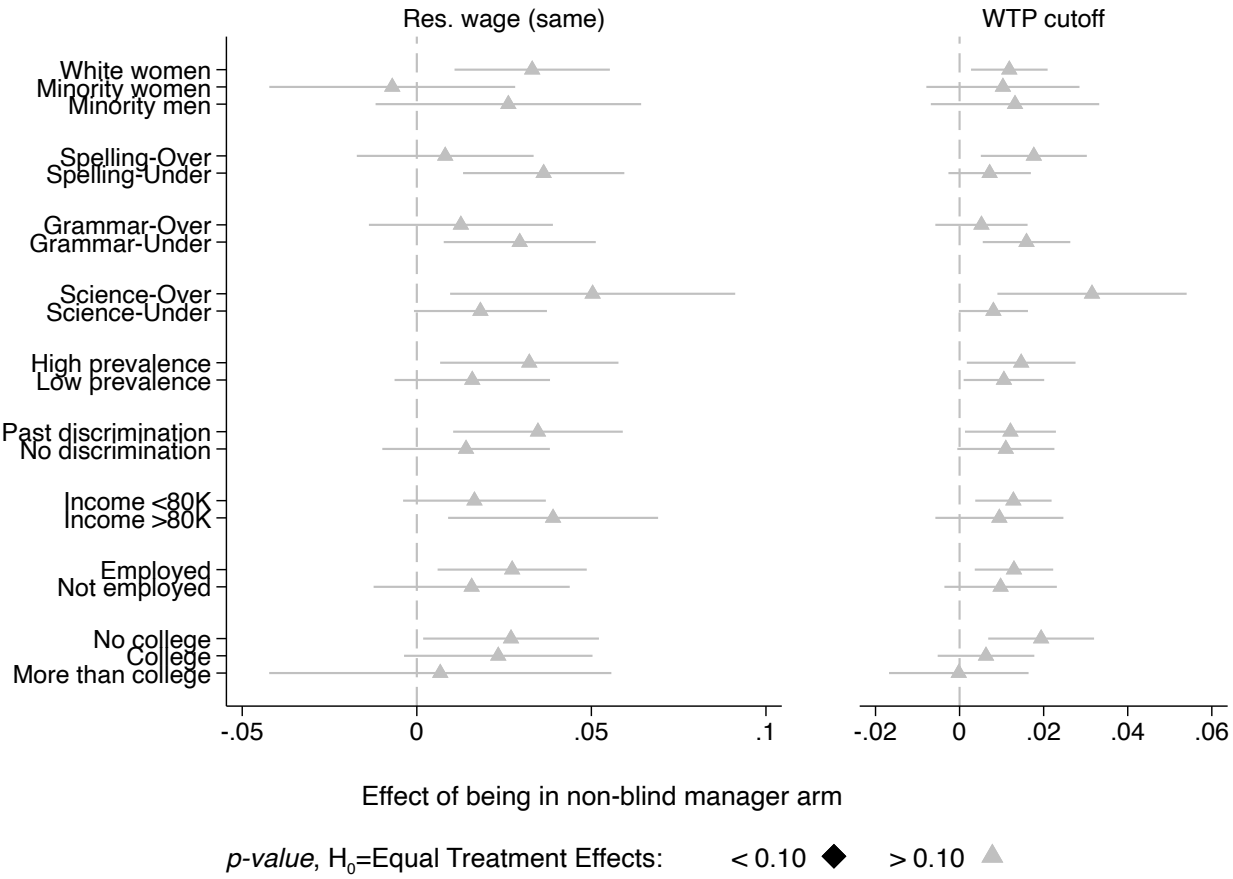
[Click to return to Section 3.3 (Effects of Perceived Discrimination)]

Figure A7: Heterogeneity by all potential mechanism variables, effort and performance outcomes



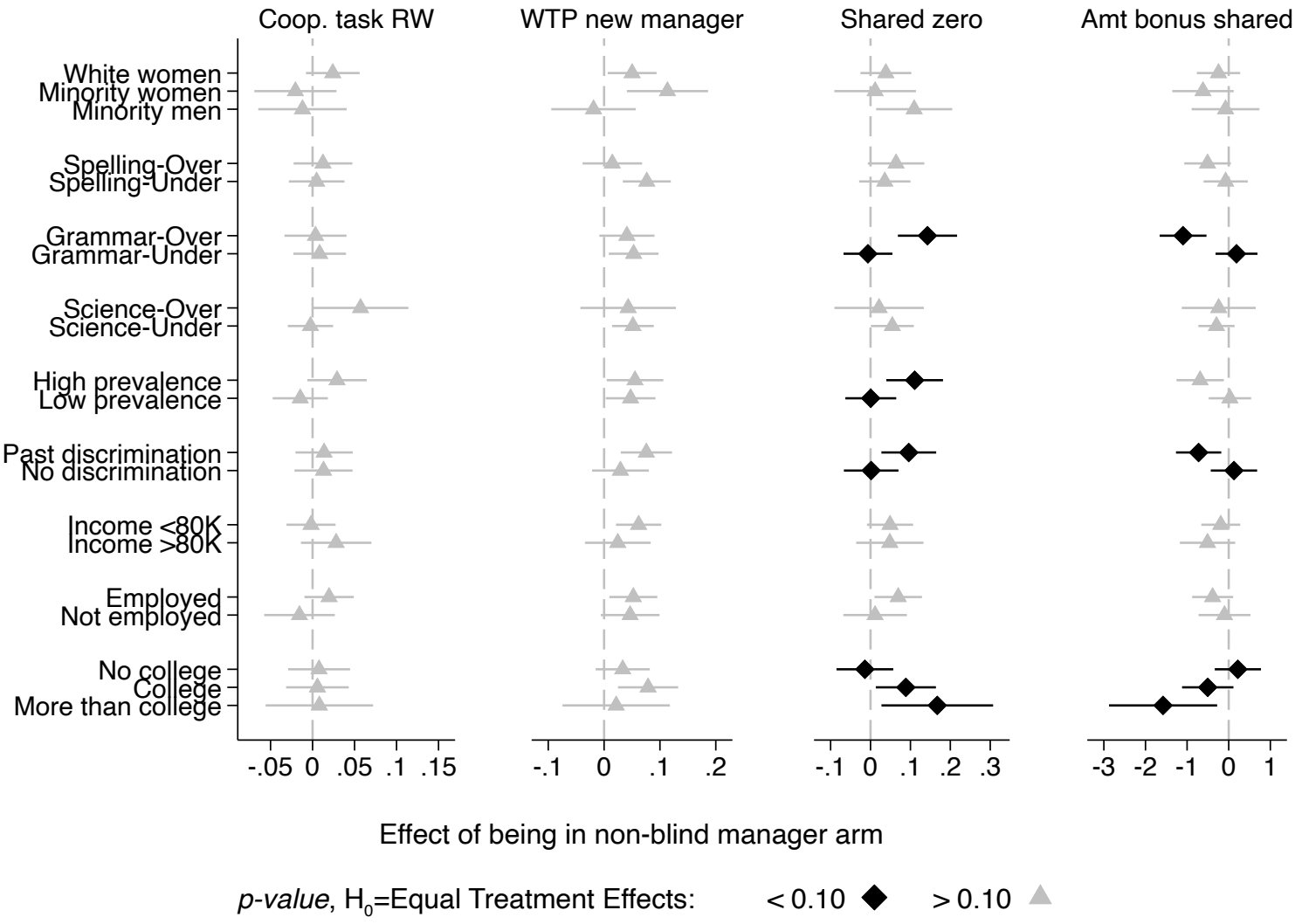
Note: This figure is analogous to Appendix Figure A6. The outcomes are the same as in Table 3.
[Click to return to Section 3.3 (Effects of Perceived Discrimination)]

Figure A8: Heterogeneity by all potential mechanism variables, future labor supply outcomes



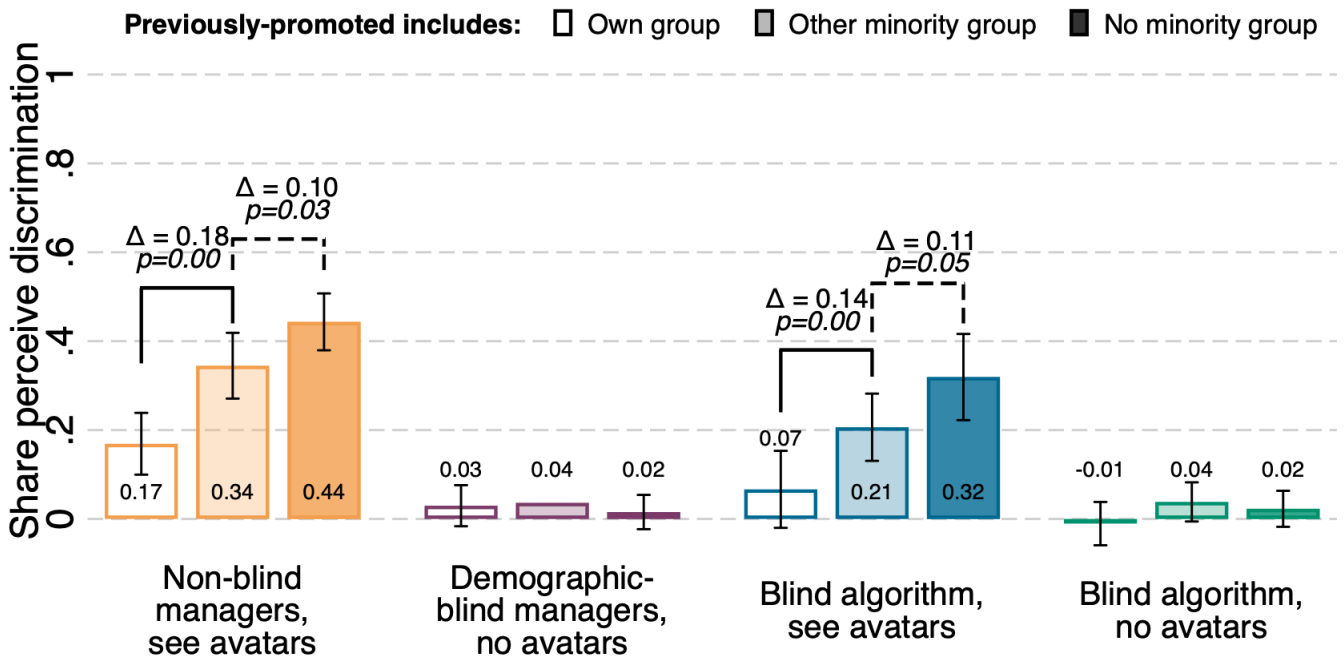
Note: This figure is analogous to Appendix Figure A6. The outcomes are the same as in Table 4.
[Click to return to Section 3.3 (Effects of Perceived Discrimination)]

Figure A9: Heterogeneity by all potential mechanism variables, cooperation and retaliation outcomes



Note: This figure is analogous to Appendix Figure A6. The outcomes are the same as in Table 5.
[Click to return to Section 3.3 (Effects of Perceived Discrimination)]

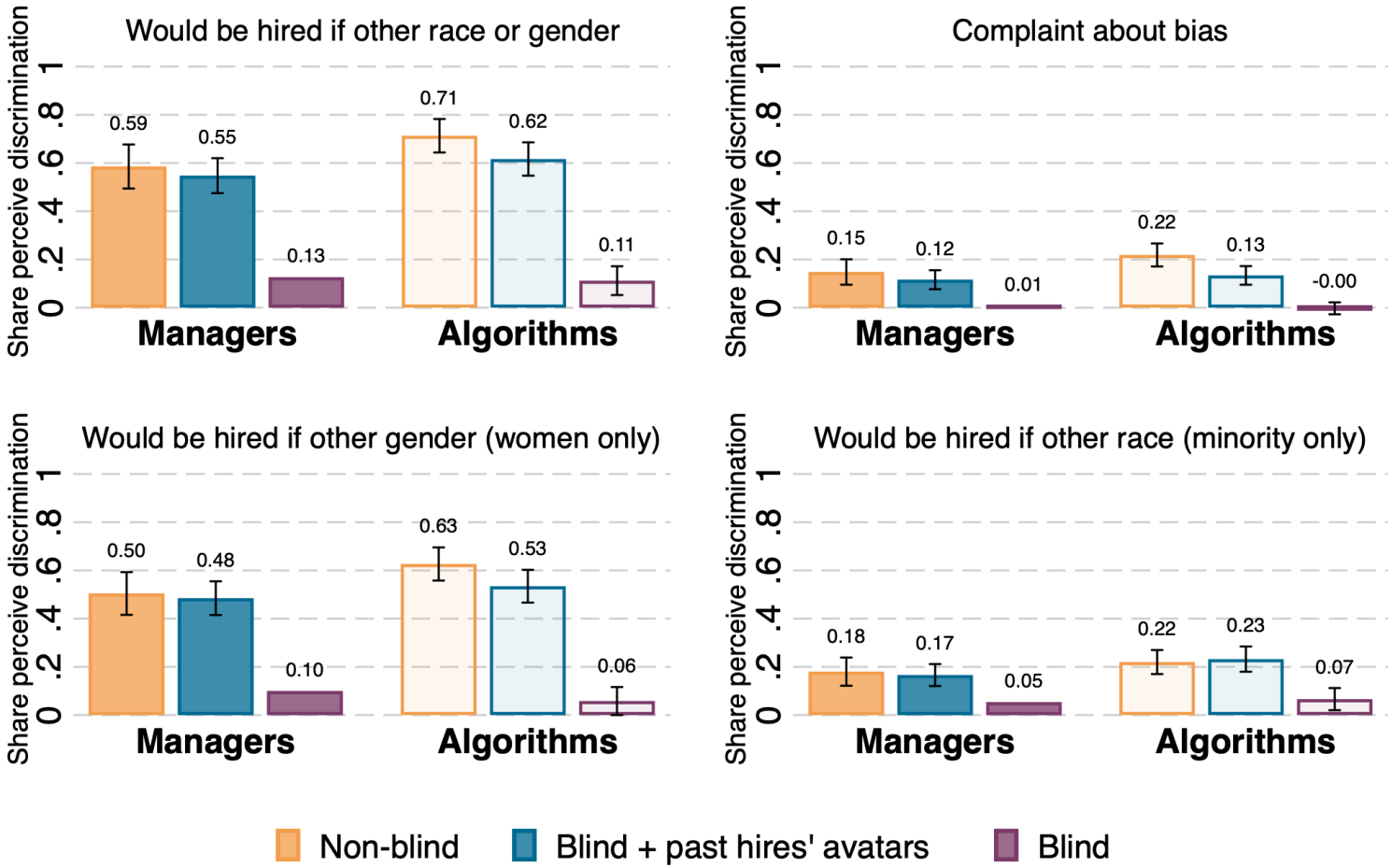
Figure A10: Effects of seeing one previously-promoted minority-group worker of a same or different demographic group



Note: This figure plots the share in each treatment arm of the promotion experiment perceiving discrimination by the main measure, separately by whether workers (would) see three white men previously promoted or not, analogous to Figure 11, but now splitting the sample who see two white men and one member of a minority group by whether the minority-group representative is of the worker's own demographic group or a different one. The sample restricts to workers who would not have been promoted under any procedure and are not white men. Controls are the same as in Figure 11. *p*-values and 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

[Click to return to Section 4 (Effects of Minority-Group Representation)]

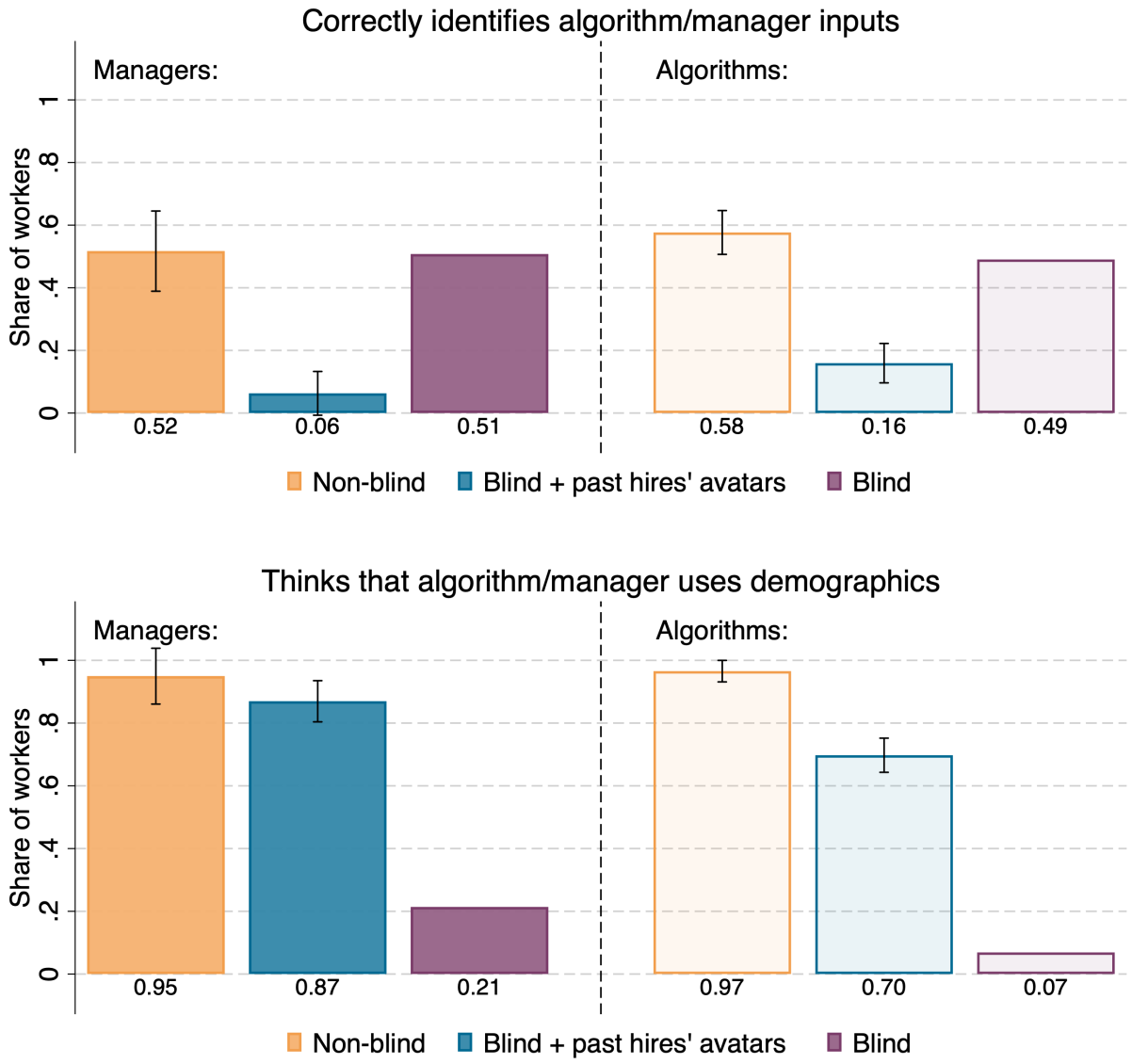
Figure A11: Perceived discrimination in the hiring sample, secondary measures of perceived discrimination



Note: This figure replicates Figure 12 using the four secondary measures of perceived discrimination. In the top graphs, the sample is restricted to workers who would not have been hired under any procedure and those who saw that three white men previously promoted. In the bottom graphs, the sample is further restricted to women only when perceived gender discrimination is the outcome and racial minorities only when perceived racial discrimination is the outcome. The secondary measures of perceived discrimination are the same as in Appendix Figure A2. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

[Click to return to Section 5 (Effects of Hiring Policies)]

Figure A12: Comprehension of hiring procedure inputs



Note: This figure shows worker knowledge of the decision-making processes by treatment arm in the hiring experiment. In the top panel, the y-axis is the share of workers correctly identifying the inputs (what decision-makers know about workers) to hiring decision-makers. When asked what information their manager or the algorithm had about workers when deciding who to hire, this requires checking off "An avatar," "Gender," "Race/Ethnicity," "1-5 stars representing the average quiz scores," and "Education" (in the non-blind group), and just quiz scores and education in the demographic-blind groups, and **not** checking off "Age," "All three numeric quiz scores from the baseline survey (science, spelling, and grammar)," "The average quiz score (for example, 85%)," "Work history," and "Time on Prolific." In the bottom panel, it is the share of workers who think that the decision-maker knew either their race and gender (i.e., checked either of those boxes on the same question as just described). The sample is restricted to workers who would not have been promoted under any hiring procedure, and workers who saw that all three previous hires were white men (a random subsample within each treatment arm—results are the same for the full sample). Shares are calculated via the same regressions as Figure 12. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

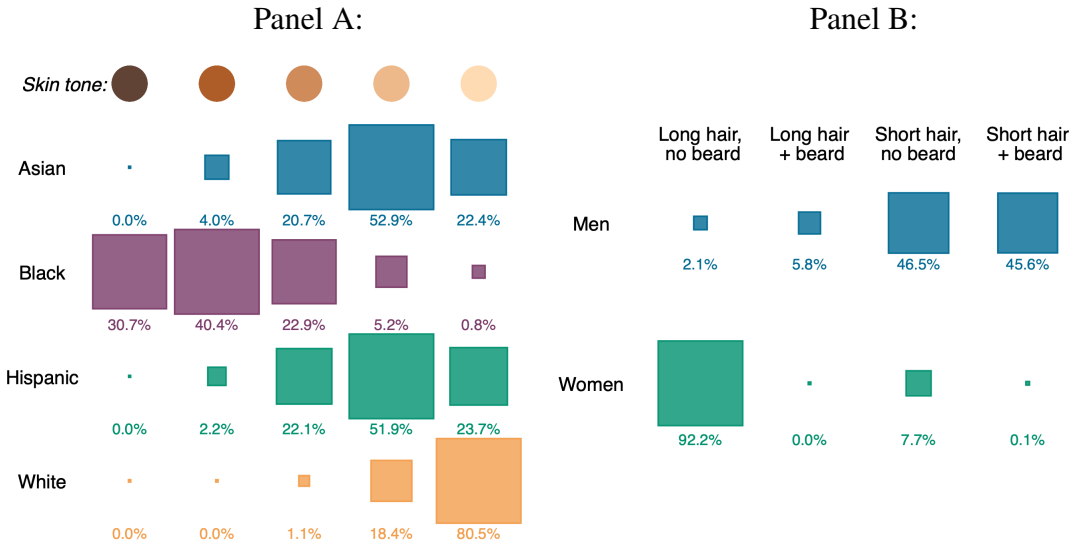
[Click to return to Section 5 (Effects of Hiring Policies)]

Figure A13: Avatar-making procedure

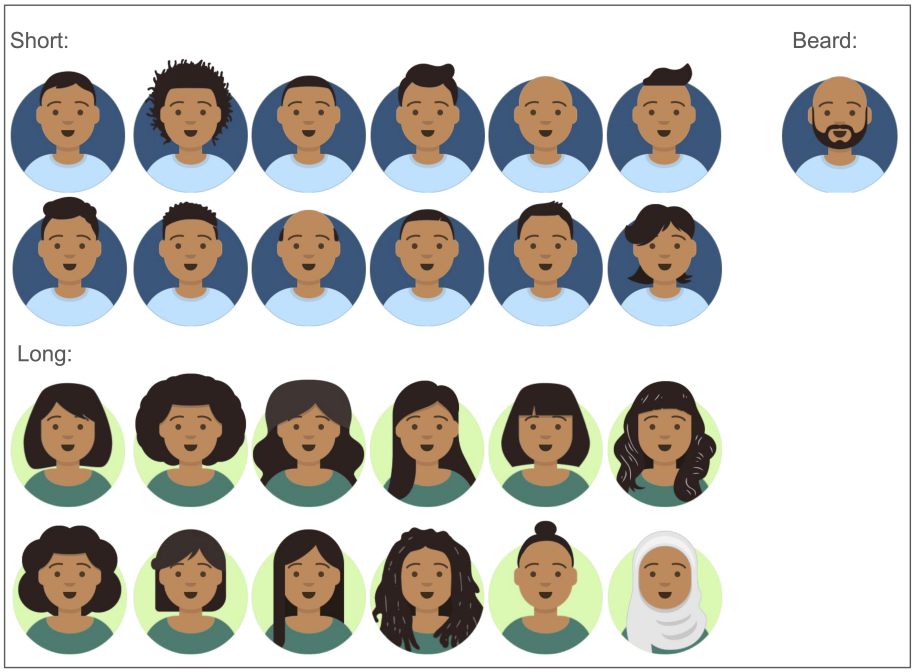


Note: This figure depicts the procedure by which workers created their avatars in the baseline survey.
[Click to return to Section 2 (Research Design)]

Figure A14: Avatar characteristics by self-reported race and gender



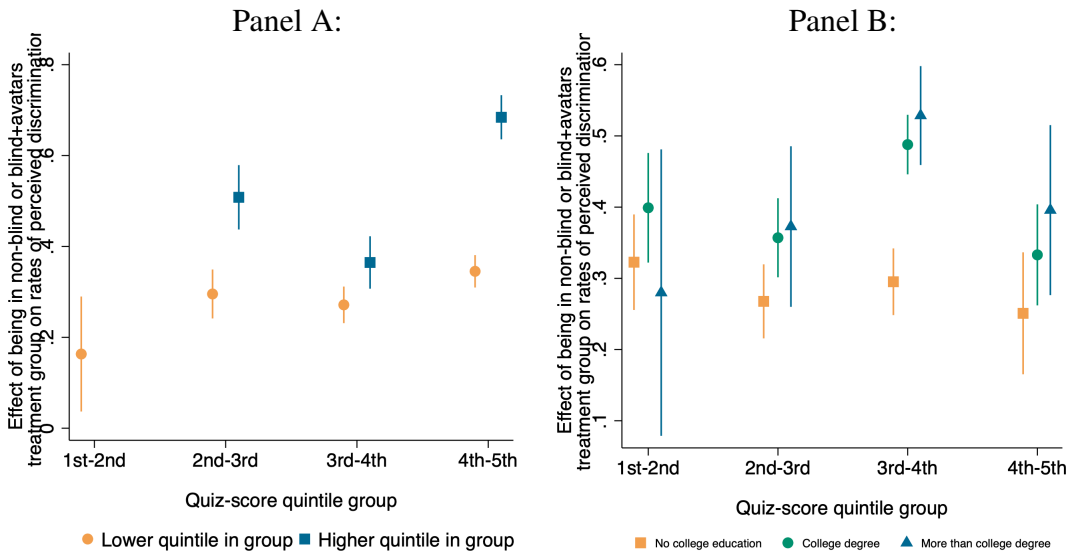
Panel C:



Note: The sample in panels A and B is all workers in the hiring and promotion experiments who took the baseline survey (N=6720). Each row plots the share of participants in that racial or gender group who choose each skin tone (Panel A) or hair style (Panel B). Panel C shows the hair styles by whether they are "long" or "short" and the representation of a beard; men could choose to add a beard to any hairstyle. Men could choose long hair with blue shirts and women could choose short hair with green shirts; these are not shown for the sake of space.

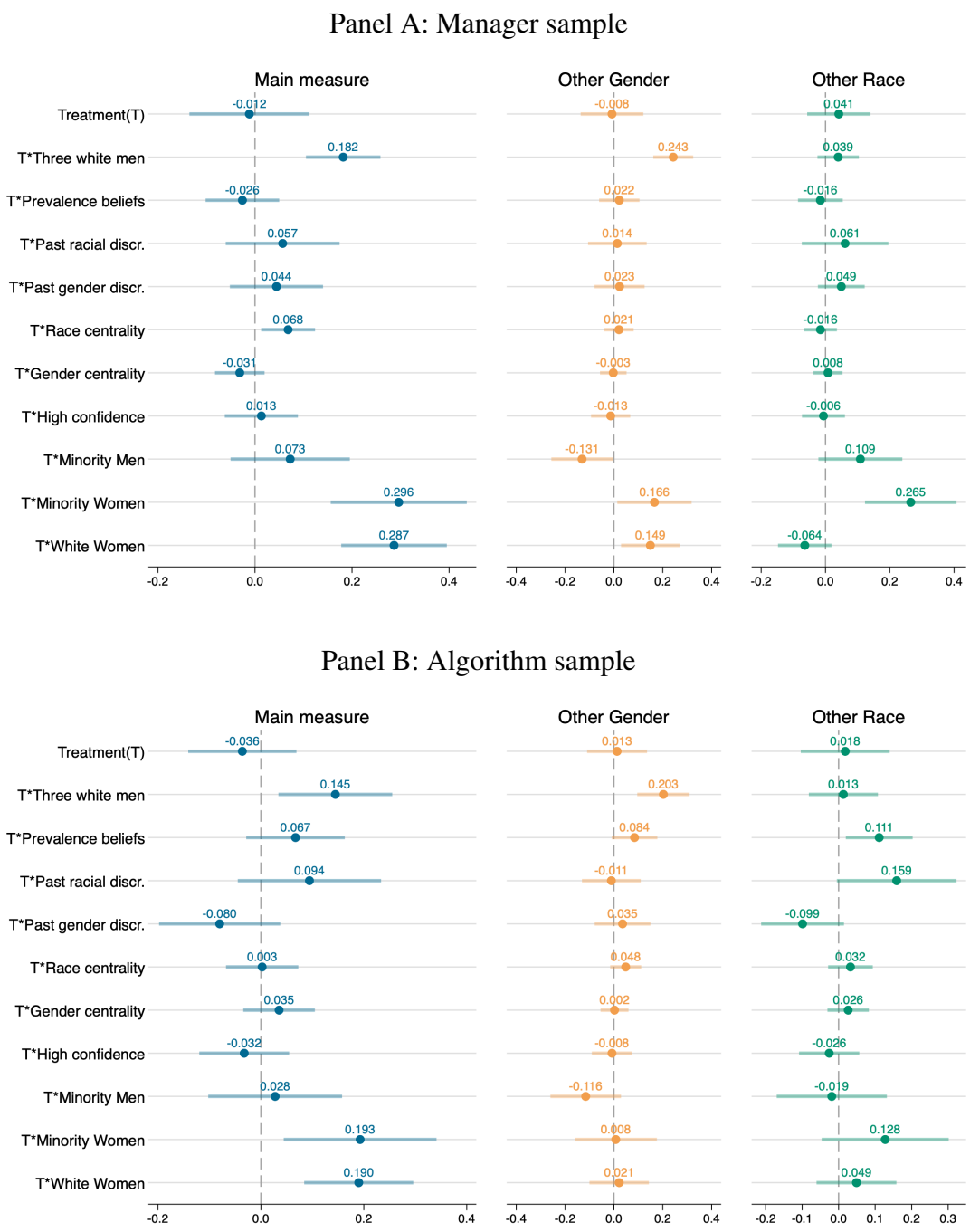
[Click to return to Section 2 (Research Design)]

Figure A15: Perceived discrimination by quiz-score quintile group and quintile



Note: The sample is all workers in the hiring and promotion experiments who would not have been hired/promoted by any of the relevant mechanisms and are not white men, and the "treatment effect" pools the effect of being in any of the arms with positive perceived discrimination in Figures 6, 7 and 12, relative to the arms with no perceived discrimination (the arms with demographic-blind decision-makers and no demographics of previously-selected workers shown); the experiments and treatment arms are pooled for power and ease of exposition but the result is very similar when looking at the non-blind manager arm in the promotion experiment separately (the other arms are too small to consider separately). In each panel, the treatment effects on perceived discrimination by quiz-score quintile group and whether the worker is in the higher quintile in their group or the lower quintile in their group (panel A) or whether they have no degree, a college degree, or more than a college degree (panel B) are jointly estimated in one regression with all of the main analysis control variables. 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 2 (Research Design)]

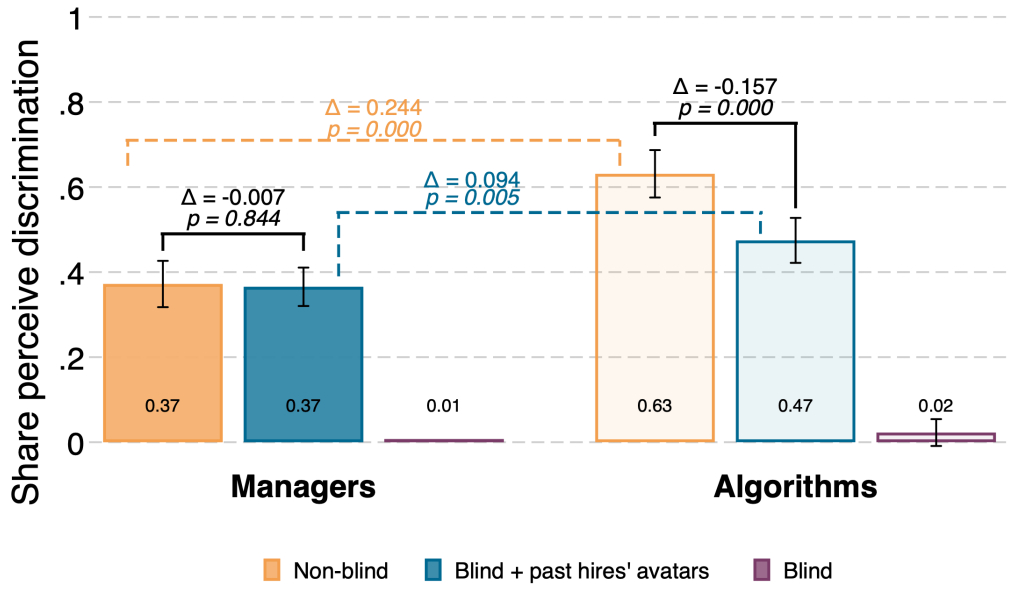
Figure A16: All predictors of perceived discrimination in the promotion experiment



Note: This figure plots coefficients from a regression of the primary and secondary measures of perceived discrimination on an indicator for treatment, that indicator interacted with each of the worker characteristics listed on the left, the worker characteristics, and the control variables in the main specification. The sample is workers who would not have been promoted under any procedure in the promotion experiment. In Panel A, *Treatment* is assignment to the non-blind manager arm among those evaluated by a manager. In Panel B, it is assignment to see previously-promoted workers' avatars, gender, and race among those evaluated by the algorithm. 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

[Click to return to Section 2 (Research Design)]
[Click to return to Section 3 (Effects of Perceived Discrimination)]

Figure A17: Perceived discrimination in the hiring experiment (full sample)



Note: This figure replicates Figure 12 but for all workers regardless of the demographic makeup of the previously-hired workers they saw, without accounting for differences across hiring procedures in the rate at which workers saw that all three previous hires were three white men versus seeing two white men and someone else. It plots the share of workers perceiving discrimination in each treatment arm of the hiring experiment, using the main measure of perceived discrimination. The sample is restricted to workers who would not have been promoted under any hiring procedure. p-values and 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 5 (Effects of Hiring Policies)]

Table A1: Correlates of perceived discrimination in the literature (labor market outcomes)

		Perceived discrimination...		
Paper	**Sample**	**Measurement**	**Frequency**	**Correlates (direction)**
Sanchez and Brock 1996	N=139 Hispanic male and female employees in Florida from multiple organizations	Agreement on a 1-5 scale with 10 items about ethnicity as a limitation at work, stereotyping, and exclusion due to ethnicity or other differences	Not reported (NR)	work tension (+), job satisfaction (-), organizational commitment (-)
Gutek et al. 1996	205 female PhD psychologists employed in organizations; 45 female senior managers taking an executive education class	Agreement on 1-4 scale with: men are promoted faster than women, my organization prefers to hire men, and men are more likely to receive tenure or its equivalent	NR	Psychologists and managers: Job involvement (0), power and prestige of current job (-), work conflict (+), likelihood of choosing same career again (-), work hours/week (+), considered changing jobs (0).
Mays et al. 1996	N=1,023 Black women of working age from a national probability sample of the adult Black population in 1979	14-item scale with four sub-indices measuring race discrimination in employment, resource or opportunity barriers, general perceived systemic discrimination, and perceived past and future racial discrimination	NR	labor supply (0), job stress (+)

Appendix Table A1 Continued

Paper	Sample	Perceived discrimination...		
		Measurement	**Frequency**	**Correlates (direction)**
Johnson and Neumark 1997	N=3,001 men from national probability sample of men aged 45-59 in 1966, observed 1971, 76, and 80	During the past five years, do you feel you were discriminated against at work because of your age?	7%	Job separations (+), non-employment spells (+)
Ensher et al. 2001	N=366 non-professional employees in Los Angeles area; 70% women; 45% Latino, 15% African American, 10% Asian, 22% Caucasian	By coworkers: Agreement (1-4 scale) with 4 statements e.g. "my coworkers sometimes make sexist statements" By supervisors: Agreement with 4 statements e.g. "my supervisor sometimes makes racist decisions" By organization: Agreement with 3 statements e.g. "I feel that some of the polices and practices of this organization are racist/sexist"	~30% report organizational policies are sexist or racist, coworkers make racist statements, and disagree that coworkers are supportive of people from diverse backgrounds. 20% report supervisors making racist or sexist decisions.	Job satisfaction (-), organizational commitment (-), organizational citizenship behavior (conscientiousness and altruism at work), grievances/complaints (0)
Goldsmith et al. 2004	N=3,146 multi-stage probability sample in Los Angeles and Boston in 1992	Have you ever felt you were refused a job because of your race/ethnicity [gender]? During the past year were you discriminated against at your work because of your race/ethnicity? Have you ever felt that others at your workplace got promotions or pay raises faster than you because of your gender?	Percent perceive race or gender discrimination in job search or on-the-job Search: R / G; OTJ: R / G Black M 47 / 14 / 17 / 3 Black W 46 / 28 / 16 / 13 Hisp. M 20 / 4 / 8 / 14 Hisp. W 13 / 10 / 7 / 6 Asian M 18 / 1 / 3 / 1 Asian W 5 / 2 / 3 / 2 White M 12 / 6 / 3 / 1 White W 6 / 14 / 5 / 15	Job search: Weeks worked in past year (-) but only statistically significant for women perceiving racial discrimination; mixed results by race as well. OTJ: Weeks worked in past year (+) for racial discrimination, (-) for gender discrimination, mixed by race

Appendix Table A1 Continued

Paper	Sample	Perceived discrimination...		
		Measurement	**Frequency**	**Correlates (direction)**
Minnotte 2012	N=2335 nationally representative sample of working adults in 2002; 75% White, 10% Black, 10% Hispanic	Have you been discriminated against at work ... because you are a woman [man] because of your race or national origin because of your age	Sex: 8% Race: 6% Age: 7%	Work-to-life conflict (index of lack of work-life balance or time for family commitments, in bad mood at home because of work): Sex (+), Race (+), Age (+)
Stainback and Irvin 2012	Same as Minnotte 2012 with some further restrictions by age, non-self-employed (N=2062)	Have you been discriminated against at work because of your race or national origin?	Black: 17% Latino: 13% White: 4%	Employer loyalty (-, Black and Latino only), job search intentions (+)
Hahn and Wilkins 2013	N=12-13,000 nationally representative sample of Australian adults in 2008-2010	In job applications in the last two years have you ever been unsuccessful because the employer discriminated against you? If yes, what attribute? In paid jobs in last two years, do you feel your employer in any way discriminated against you because of your gender, age, ethnicity, religion, or parenting responsibilities?	8.5% of job applicants 7.5% of employees More common for older respondents, women, women with young children (in-job only)	In-job perceived discrimination: Wage levels (0), self-assessed over-employment (+) and under-employment (+), wage changes (+), promotion (0), job satisfaction (-), self-assessed probability of job loss (+), life satisfaction (-)

Appendix Table A1 Continued

Paper	Sample	Perceived discrimination...		
		Measurement	**Frequency**	**Correlates (direction)**
Mukerjee 2014	N=2,099 from General Social Survey cross-sectional national samples in 2002, 2006, and 2010	Do you feel in any way discriminated against on your job because of your race or ethnic origin?	NR	Job satisfaction (-)
Triana et al. 2015	Meta-analysis with 39 studies	Perceived racial discrimination	N/A	Job attitudes (-) psychological health (-) physical health (-), perceived diversity climate (-), organizational citizenship behaviors (-), coping behaviors (+)
Pager and Pedulla 2015	N=1,405 national probability sample of unemployed job seekers in 2013	During the last year you worked, did you witness any discriminatory comments or actions at your workplace related to race or ethnicity? Have you felt at any time in the past that your work opportunities have been limited by your race or ethnicity?	Black: 37% yes to either; 21% witness discrimination and 33% feel opportunities limited by race White: 18% yes to either, 12% witness discrimination and 11% feel opportunities limited by race	Job search breadth (+)

Note: This is not an exhaustive list of related papers. I focus on the most well-cited studies of perceived discrimination and work outcomes that use measures of perceived discrimination that are most analogous to mine.
[Click to return to Section 1 (Introduction) or Section 3.4.2 (Mechanisms)]

Table A2: Correlates of perceived discrimination in the literature (mental health outcomes)

		Perceived discrimination...		
Paper	**Sample**	**Measurement**	**Frequency**	**Correlates (direction)**
Williams et al. 1997	N=1,139 multistage area probability sample in Detroit, MI weighted to area population (1995)	Major events: Do you think you have ever been unfairly fired or denied promotion, not hired, or stopped, searched, questioned, threatened, or abused by police? Day-to-day: How frequently do you experience the following 9 types of discrimination; e.g. being treated with less courtesy or respect than others, being called names, being threatened, receiving poorer service	Any major event: 62% (Black), 36% (White) Day-to-day: not reported (NR)	Major events: No correlation with four health measures when other stressors are controlled Day-to-day: Overall health (-), bed-days (+), psychological distress (+), psychological well-being (-)
Kessler et al. 1999	N=3,032 random digit dialing sample weighted to US population (1996)	Major events: Have you ever been discriminated against in e.g. job promotion, housing, dealing with financial institutions. Day-to-day: Same as Williams et al. 1997 (never=0, ever=1)	Any major event: 34% Any day-to-day: 61%	Major events: Major Depression (MD) (+), psychological distress (+) Day-to-day: MD (+), Generalized Anxiety Disorder (GAD) (+), psychological distress (+)

Appendix Table A2 Continued

Paper	Sample	Perceived discrimination...		
		Measurement	**Frequency**	**Correlates (direction)**
Pascoe and Smart Richman 2009	Meta-analysis from 134 samples from 1986-2007	Most common: Racial discrimination: Perceived Racism Scale, Schedule of Racist Events, Index of Race Related Stress, Racism and Life Experiences Scale; Gender Discrimination: Schedule of Sexist Events; Unspecified: Williams et al.'s 1997 Scale; all generally elicited in various domains	N/A	Mental health (-), physical health (-), participation in healthy behaviors (-)
Lee and Turney 2012	N=3,105 multistage probability sample in Chicago stratified by neighborhood (2001-2003)	Same as Kessler et al. 1999	NR	Major events: Depressive symptoms (0), loneliness (0), hostility (+); Day-to-day: Depressive symptoms (+), loneliness (+), hostility (+)
Schmitt et al. 2014	Meta-analysis N=144,246 from 328 samples	Personal discrimination (296 studies): How frequently or whether certain discriminatory events have happened Group discrimination: Group X is often confronted with discrimination (55 studies)	N/A	Personal and group discrimination: psychological well-being (-). More negative correlation for negatively-framed psychological outcomes (e.g. depression, distress, etc.) than positively-framed outcomes (e.g. self-esteem)

Appendix Table A2 Continued

Paper	Sample	Perceived discrimination...		
		Measurement	**Frequency**	**Correlates (direction)**
Everett et al. 2016	N=14,609 nationally representative sample in young adulthood (AddHealth Wave IV, 2007-2008)	In your day-to-day life, how often do you feel you are treated with less respect or courtesy than other people?	24% (never/rarely=0, sometimes/often=1)	Perceived stress (+), depressive symptoms (+), anxiety (0)
Hackett et al. 2020	N=4,883 ethnic minority participants in nationally representative UK sample (2009-2010)	Have you (a) felt unsafe, (b) avoided going to or being in, (c) been insulted, called names, threatened or shouted at, or (d) been physically attacked in 7 different settings? Then attribute to e.g. ethnicity, age, sex, etc.	20% (never any racial discrimination=0, ever any racial discrimination=1)	Psychological distress (+), mental functioning (-), life satisfaction (-), limiting longstanding illness (+), physical functioning (-), poor self-rated health (+)

Note: This is not an exhaustive list of related papers. I focus on the largest-sample studies of perceived discrimination and mental health using data collected in the last 30 years and meta-analyses. Many other studies cited in these papers provide key qualitative evidence among specific communities but are not included here.
[Click to return to Section 1 (Introduction) or Section 3.4.2 (Mechanisms)]

Table A3: Balance table, promotion experiment

	Manager arms			Algorithm sub-arms		
	Blind	Non-blind	*p-value*	No avatars	Avatars	*p-value*
	(1)	(2)	(1)=(2)	(3)	(4)	(3)=(4)
Male	0.28	0.27	*0.84*	0.35	0.31	*0.25*
Race:						
Asian	0.13	0.12	*0.62*	0.14	0.16	*0.42*
Black	0.13	0.13	*0.93*	0.17	0.13	*0.19*
White, Hispanic	0.15	0.14	*0.64*	0.14	0.13	*0.66*
White, Non-Hispanic	0.59	0.61	*0.46*	0.55	0.58	*0.50*
Married	0.54	0.52	*0.37*	0.56	0.54	*0.60*
Kids	0.43	0.45	*0.46*	0.42	0.42	*0.85*
Education:						
Less than high school	0.01	0.01	*0.40*	0.01	0.01	*0.91*
High school graduate	0.14	0.15	*0.65*	0.11	0.13	*0.35*
Some college but no degree	0.29	0.27	*0.36*	0.25	0.27	*0.54*
2 year college degree	0.10	0.10	*0.78*	0.10	0.07	*0.23*
4 year college degree	0.36	0.34	*0.39*	0.39	0.41	*0.62*
Professional or Masters degree	0.09	0.13	*0.03*	0.13	0.10	*0.15*
Doctorate	0.01	0.02	*0.13*	0.01	0.01	*0.85*
Income:						
Less than $20,000	0.13	0.14	*0.44*	0.15	0.14	*0.78*
$20,000-$40,000	0.24	0.21	*0.14*	0.20	0.23	*0.40*
$40,000-$60,000	0.17	0.20	*0.19*	0.16	0.20	*0.17*
$60,000-$80,000	0.15	0.15	*0.83*	0.14	0.16	*0.49*
$80,000-$100,000	0.12	0.11	*0.54*	0.10	0.08	*0.49*
$100,000-$120,000	0.07	0.07	*0.91*	0.07	0.05	*0.29*
$120,000-$140,000	0.04	0.02	*0.05*	0.06	0.04	*0.23*
$140,000-$160,000	0.03	0.04	*0.16*	0.04	0.03	*0.41*
More than $160,000	0.05	0.05	*0.72*	0.08	0.07	*0.55*
Age:						
18-24	0.17	0.16	*0.71*	0.17	0.18	*0.76*
25-34	0.31	0.32	*0.70*	0.32	0.33	*0.78*
35-44	0.22	0.22	*0.78*	0.20	0.23	*0.36*
45-54	0.15	0.13	*0.43*	0.15	0.12	*0.25*
55-64	0.11	0.11	*0.87*	0.12	0.08	*0.14*
65 or older	0.04	0.06	*0.18*	0.04	0.06	*0.34*
Employment:						
Currently employed outside Prolific	0.66	0.65	*0.67*	0.67	0.72	*0.23*
Ever employed outside Prolific	0.96	0.95	*0.36*	0.95	0.97	*0.14*
Satisfied with current employer	0.78	0.79	*0.66*	0.77	0.75	*0.67*
Satisfied with most recent employer	0.66	0.62	*0.38*	0.64	0.65	*0.93*
Past experienced discrimination:						
Job search	0.26	0.24	*0.23*	0.28	0.24	*0.27*
Promotion	0.19	0.18	*0.40*	0.18	0.14	*0.15*
Termination	0.12	0.09	*0.09*	0.09	0.09	*0.85*
Daily work activities	0.34	0.32	*0.30*	0.36	0.28	*0.04*
Any discrimination	0.52	0.50	*0.62*	0.53	0.46	*0.05*
Beliefs about prevalence (own group)	0.47	0.44	*0.05*	0.46	0.45	*0.74*
N	694	695		371	320	

Note: This table shows means of worker demographic characteristics, all measured in the baseline survey, by treatment group in the promotion experiment and tests for balance across the manager arms and the algorithm sub-arms. The sample is workers who would not have been promoted by any of the three procedures. All variables except for the final one are indicators; the final variable is a continuous, self-reported belief about the share of workers in one's own race $\times$ gender group who report experiencing discrimination in the past in the baseline survey. *p*-values are calculated with standard errors robust to heteroskedasticity.

[Click to return to Section 2 (Research Design)]

Table A4: Balance table, hiring experiment

	Manager arms			Algorithm arms			
	Blind	Blind+	Non-Blind	Blind	Blind+	Non-Blind	*joint*
	(1)	(2)	(3)	(4)	(5)	(6)	*p-value*
Male	0.12	0.16	0.17	0.15	0.13	0.11	*0.55*
Race:							
Asian	0.06	0.06	0.09	0.07	0.06	0.05	*0.61*
Black	0.13	0.13	0.10	0.12	0.13	0.12	*0.69*
White, Hispanic	0.09	0.10	0.10	0.12	0.11	0.12	*0.44*
White, Non-Hispanic	0.73	0.71	0.71	0.69	0.70	0.71	*0.66*
Married	0.60	0.61	0.53	0.55	0.54	0.61	*0.23*
Kids	0.45	0.46	0.39	0.44	0.45	0.50	*0.52*
Education:							
Less than high school	0.01	0.00	0.01	0.00	0.01	0.00	*0.99*
High school graduate	0.10	0.09	0.08	0.08	0.12	0.09	*0.20*
Some college but no degree	0.29	0.27	0.26	0.25	0.24	0.23	*0.53*
2 year college degree	0.10	0.12	0.13	0.12	0.10	0.08	*0.52*
4 year college degree	0.36	0.37	0.34	0.38	0.37	0.44	*0.14*
Professional or Masters degree	0.13	0.13	0.16	0.15	0.14	0.14	*0.76*
Doctorate	0.02	0.01	0.02	0.02	0.03	0.01	*0.14*
Income:							
Less than $20,000	0.11	0.10	0.10	0.11	0.12	0.10	*0.85*
$20,000-$40,000	0.18	0.17	0.21	0.16	0.18	0.15	*0.59*
$40,000-$60,000	0.19	0.25	0.18	0.19	0.16	0.19	*0.28*
$60,000-$80,000	0.17	0.14	0.15	0.16	0.17	0.18	*0.92*
$80,000-$100,000	0.15	0.09	0.07	0.11	0.13	0.14	*0.19*
$100,000-$120,000	0.07	0.07	0.11	0.08	0.07	0.07	*0.26*
$120,000-$140,000	0.06	0.07	0.08	0.07	0.05	0.06	*0.96*
$140,000-$160,000	0.03	0.02	0.04	0.04	0.05	0.04	*0.55*
More than $160,000	0.05	0.08	0.07	0.07	0.07	0.07	*0.44*
Age:							
18-24	0.17	0.13	0.17	0.14	0.15	0.13	*0.61*
25-34	0.31	0.33	0.28	0.34	0.33	0.35	*0.61*
35-44	0.20	0.25	0.22	0.22	0.21	0.22	*0.86*
45-54	0.19	0.17	0.19	0.17	0.14	0.16	*0.33*
55-64	0.09	0.10	0.09	0.10	0.10	0.12	*0.89*
65 or older	0.03	0.02	0.05	0.04	0.06	0.02	*0.16*
Employment:							
Currently employed outside Prolific	0.66	0.72	0.70	0.70	0.67	0.72	*0.45*
Ever employed outside Prolific	0.97	0.98	0.97	0.97	0.97	0.97	*0.99*
N	411	449	389	420	443	415	

Note: This table shows means of worker demographic characteristics, all measured in the baseline survey, by treatment group in the hiring experiment. The sample is workers who would not have been hired under any of the four procedures. *p*-values test the null of equality across all groups, and are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 2 (Research Design)]

Table A5: Share of workers holding negative first- or second-order stereotypes

	Share expect negative stereotypes against...			
	Women	Asian	Black	Hispanic
	(1)	(2)	(3)	(4)
First-order stereotype	0.521	0.110	0.497	0.598
Second-order stereotype	0.747	0.089	0.557	0.467
First- or second-order stereotype	0.824	0.149	0.664	0.685
First- and second-order stereotype	0.443	0.051	0.390	0.381

Note: This table shows the share of participants who themselves hold negative stereotypes against each demographic group (row 1), believe white men hold negative stereotypes against each demographic group (row 2), and those who report either of these (row 3). A "negative stereotype belief" indicates believing that the group in the column performs worse than the comparison group by any amount (row 1) or believing that white men think so (row 2). In column 1, the comparison group is men; in columns 2-4, it is White, Non-Hispanic workers. The sample is participants with both first- and second-order beliefs (N=336).

Table A6: Effects on secondary measures of perceptions of discrimination in the manager sample

	All Workers	Women	Minority Men	White Men
Panel A: Free response, primary reason is discrimination				
Non-blind manager, see avatars	0.181***	0.194***	0.168***	0.078*
	(0.017)	(0.021)	(0.033)	(0.045)
Control mean	0.010	0.014	0.000	0.000
Panel B: Yes, would be promoted if different race or gender				
Non-blind manager, see avatars	0.254***	0.290***	0.163***	0.086
	(0.024)	(0.029)	(0.056)	(0.056)
Control mean	0.127	0.130	0.159	0.000
Panel C: Complaint about job assignment mentions discrimination				
Non-blind manager, see avatars	0.100***	0.117***	0.050**	0.048
	(0.013)	(0.016)	(0.022)	(0.050)
Control mean	0.003	0.004	0.000	0.000
*Panel D: Members of my race * gender group underrepresented among promoted*				
Non-blind manager, see avatars	0.165***	0.213***	-0.005	0.007
	(0.027)	(0.032)	(0.064)	(0.047)
Control mean	0.386	0.378	0.538	0.020
Panel E: White men are overrepresented among promoted				
Non-blind manager, see avatars	0.161***	0.196***	0.022	0.327**
	(0.026)	(0.030)	(0.062)	(0.143)
Control mean	0.585	0.580	0.614	0.551
N	1389	1004	291	94

Note: This table shows the effects on secondary measures of perceived discrimination of being in the non-blind manager arm, relative to the demographic-blind manager arm, in the promotion experiment. The sample is workers in either manager arm who would not have been promoted under any of the three procedures. The secondary measures are the following: In Panel A, an indicator for the worker's *primary* reason listed for what needed to be different about their profile being demographics (as opposed to the main measure, which indicates mentioning demographics at all); in Panel B, an indicator for saying they think they would have been promoted if their race or gender was different; and in Panel C, an indicator for mentioning bias or discrimination in their open-ended description after saying they had a complaint about the promotion procedure. These are alternative measures of discrimination against the person themself. Panels D and E measure perceptions of discrimination in general: an indicator for thinking that people in their race $\times$ gender group are under-represented among the promoted workers, and an indicator for thinking that white men are over-represented among the promoted workers. Specifications are otherwise the same as in Figure 6. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.

[Click to return to Section 3.2 (Effects of Perceived Discrimination)]

Table A7: Race and gender gaps in confidence, discrimination prevalence and past experience, outside options, and competitiveness

	White women (1)	Minority women (2)	Minority men (3)	*p-value (1)=(2)*	*p-value (2)=(3)*
Variables observed in promotion experiment only					
Ever experienced discrimination at work	0.497	0.506	0.416	0.753	0.011
Discrimination prevalence (own group)	0.407	0.608	0.517	0.000	0.000
N	1052	416	442		
Variables observed in both experiments					
Currently employed	0.677	0.661	0.714	0.377	0.022
No college degree	0.372	0.396	0.411	0.208	0.555
4-year degree	0.471	0.494	0.478	0.246	0.516
More than 4-year degree	0.157	0.109	0.111	0.001	0.909
Earn more than 80k	0.348	0.307	0.342	0.029	0.132
Average percent score on quizzes	72.391	69.341	68.375	0.000	0.217
Over-confident, spelling	0.436	0.371	0.455	0.001	0.001
Over-confident, grammar	0.422	0.384	0.435	0.052	0.037
Over-confident, science	0.143	0.204	0.274	0.000	0.001
N	2858	805	774		
Variables observed in hiring experiment only					
Competitiveness (these tasks)	6.565	6.622	7.322	0.653	0.000
Above-median competitiveness	0.395	0.416	0.503	0.429	0.020
N	1806	389	332		

Note: This table compares potential mechanism characteristics across race and gender. The first row shows the share of workers who report experiencing discrimination at work in the past. The second row shows the average of workers' beliefs about the share of workers who report experiencing discrimination at work in the past in their own demographic group. The next five rows show the education, income, employment, and baseline quiz performance of each group of workers. The over-confidence variables in the subsequent three rows are indicators for whether a participant thinks they got more questions right on the corresponding quiz than they actually did; shares are shown. The final two variables show the average self-reported competitiveness in tasks like these (the spelling, science, and grammar quiz) on a scale from 0 to 10, with 10 being very competitive and 1 being not competitive at all, and an indicator for having above-median (7) competitiveness.

Table A8: Effects on retention, gender and racial heterogeneity

	Num paragraphs	Did any optional	Did all paragraphs
	(1)	(2)	(3)
Panel A: Race and gender heterogeneity			
Non-blind manager*White women	-0.646**	-0.039*	-0.047
	(0.317)	(0.022)	(0.032)
Non-blind manager*Minority women	-0.945*	-0.025	-0.116**
	(0.553)	(0.037)	(0.057)
Non-blind manager*Minority men	0.889	0.068	0.073
	(0.562)	(0.042)	(0.054)
p-values:			
White=minority, women	*0.628*	*0.727*	*0.277*
Men=women, minority	*0.018*	*0.096*	*0.014*
N	1293	1293	1293
Panel B: Gender heterogeneity			
Non-blind manager*Women	-0.726**	-0.035*	-0.065**
	(0.284)	(0.019)	(0.029)
Non-blind manager*Minority men	0.889	0.068	0.073
	(0.562)	(0.042)	(0.054)
p-values:			
Men=women	*0.008*	*0.023*	*0.020*
N	1293	1293	1293
Panel C: Racial heterogeneity			
Non-blind manager*White women	-0.643**	-0.039*	-0.047
	(0.317)	(0.022)	(0.032)
Non-blind manager*Minority	0.004	0.023	-0.018
	(0.403)	(0.029)	(0.040)
p-values:			
White=minority	*0.186*	*0.067*	*0.564*
N	1293	1293	1293
Control Mean, white women	16.402	0.944	0.818
Control Mean, minority women	15.829	0.921	0.764
Control Mean, minority men	14.538	0.821	0.662

Note: This table estimates the effect of being in the non-blind manager arm on retention relative to the demographic-blind manager arm in the promotion experiment, separately for white women, racial minority women, and racial minority men. The outcomes and sample are the same as in Panel A of Table 2, but white men are additionally dropped from the sample. The specification is also the same, but the indicator for treatment is replaced with three interactions between treatment and indicators for being a white women, an Asian, Black, or Hispanic woman, and an Asian, Black, or Hispanic man. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

Table A9: Effects on effort and performance, gender and racial heterogeneity

	Minutes	Clicks	Correct	Incorrect	Bonus
	(1)	(2)	(3)	(4)	(5)
Panel A: Race and gender heterogeneity					
Non-blind manager*White women	-0.059	1.007	-0.038	0.092	-0.017
	(0.073)	(2.393)	(0.297)	(0.257)	(0.028)
Non-blind manager*Minority women	-0.011	4.696	0.866	0.794*	0.047
	(0.116)	(4.439)	(0.529)	(0.406)	(0.050)
Non-blind manager*Minority men	0.088	-4.750*	1.478***	-0.111	0.094**
	(0.120)	(2.638)	(0.475)	(0.487)	(0.045)
p-values:					
White=minority, women	*0.721*	*0.466*	*0.125*	*0.122*	*0.258*
Men=women, minority	*0.547*	*0.060*	*0.383*	*0.155*	*0.479*
N	1295	1295	1295	1295	1295
Panel B: Gender heterogeneity					
Non-blind manager*Women	-0.046	1.992	0.203	0.279	0.000
	(0.063)	(2.107)	(0.268)	(0.229)	(0.025)
Non-blind manager*Minority men	0.088	-4.750*	1.478***	-0.111	0.094**
	(0.120)	(2.635)	(0.475)	(0.486)	(0.045)
p-values:					
Men=women	*0.305*	*0.037*	*0.019*	*0.465*	*0.067*
N	1295	1295	1295	1295	1295
Panel C: Racial heterogeneity					
Non-blind manager*White women	-0.059	0.995	-0.037	0.090	-0.017
	(0.073)	(2.393)	(0.296)	(0.256)	(0.028)
Non-blind manager*Minority	0.040	-0.193	1.183***	0.325	0.071**
	(0.085)	(2.622)	(0.359)	(0.320)	(0.034)
p-values:					
White=minority	*0.355*	*0.735*	*0.007*	*0.547*	*0.041*
N	1295	1295	1295	1295	1295
Control Mean, white women	4.968	36.903	15.581	3.475	0.976
Control Mean, minority women	4.932	36.179	14.686	2.707	0.905
Control Mean, minority men	4.904	33.503	12.552	3.938	0.752

Note: This table estimates the effect of being in the non-blind manager arm on effort and performance relative to the demographic-blind manager arm in the promotion experiment, separately for white women, racial minority women, and racial minority men. The outcomes and sample are the same as in Panel A of Table 3, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A8. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

[Click to return to Section 3.3 (Effects on Retention and Performance)]

Table A10: Combined heterogeneity analysis: Number of paragraphs

	(1)	(2)	(3)	(4)
Treatment	-0.646**	-1.000**	-0.861*	-0.742
	(0.317)	(0.419)	(0.515)	(0.671)
Treat*Minority women	-0.299	-0.218	-0.012	-0.033
	(0.618)	(0.619)	(0.691)	(0.692)
Treat*Minority men	1.535**	1.504**	1.382**	1.380**
	(0.627)	(0.629)	(0.657)	(0.659)
Treat*Spelling confidence	—	0.696	0.863*	0.843*
		(0.496)	(0.506)	(0.508)
Treat*Grammar confidence	—	0.074	-0.012	-0.020
		(0.488)	(0.497)	(0.499)
Treat*Science confidence	—	0.098	0.073	0.114
		(0.666)	(0.689)	(0.692)
Treat*High prevalence	—	—	0.125	0.128
			(0.531)	(0.532)
Treat*Past discrimination	—	—	-0.473	-0.462
			(0.504)	(0.506)
Treat*Income>80k	—	—	—	-0.574
				(0.535)
Treat*Employed	—	—	—	0.087
				(0.546)
Treat*College degree	—	—	—	-0.006
				(0.534)
N	1293	1293	1234	1234
p-value for null hypothesis that...				
No race/gender het.:	0.029	0.039	0.092	0.093
No gender het. (minorities):	0.018	0.028	0.089	0.086
No race het. (women):	0.628	0.725	0.986	0.962

Note: This table tests for heterogeneity along all possible explanatory dimensions; sequentially adding potential explanatory variables. An interaction of each heterogeneity variable with the treatment indicator is sequentially added to the regression; the dependent variable is the number of paragraphs proofread. The indicators are: an indicator for being a racial minority woman, an indicator for being a racial minority man, indicators for being over-confident in spelling, grammar, and science, respectively, an indicator for having above-median beliefs about the prevalence of discrimination against one's own race *times* gender group, among one's group, an indicator for reporting previously experiencing discrimination at work in the past, and indicators for having income above $80,000, being employed, and having a college degree. The "treatment" variable indicates being in the non-blind manager arm. The sample and controls are the same as in Table 2, with the main effects of the heterogeneity variables added as controls. The outcome variable is the same as in the first column of that table. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table A11: Combined heterogeneity analysis: Correct highlights in required paragraphs

	(1)	(2)	(3)	(4)
Treatment	-0.038	-0.101	-0.363	-0.632
	(0.297)	(0.396)	(0.471)	(0.598)
Treat*Minority women	0.904	0.807	1.173*	1.132*
	(0.589)	(0.595)	(0.628)	(0.625)
Treat*Minority men	1.516***	1.367**	1.650***	1.577***
	(0.559)	(0.558)	(0.583)	(0.582)
Treat*Spelling confidence	—	-0.668	-0.590	-0.600
		(0.444)	(0.455)	(0.455)
Treat*Grammar confidence	—	0.650	0.605	0.586
		(0.446)	(0.454)	(0.456)
Treat*Science confidence	—	1.061*	0.916	0.895
		(0.605)	(0.620)	(0.628)
Treat*High prevalence	—	—	-0.250	-0.229
			(0.481)	(0.480)
Treat*Past discrimination	—	—	0.558	0.546
			(0.467)	(0.469)
Treat*Income>80k	—	—	—	-0.591
				(0.493)
Treat*Employed	—	—	—	0.839*
				(0.497)
Treat*College degree	—	—	—	-0.157
				(0.493)
N	1295	1295	1236	1236
p-value for null hypothesis that...				
No race/gender het.:	0.017	0.037	0.010	0.013
No gender het. (minorities):	0.383	0.430	0.517	0.544
No race het. (women):	0.125	0.175	0.062	0.071

Note: This table tests for heterogeneity along all possible explanatory dimensions; sequentially adding potential explanatory variables. The table is analogous to Appendix Table A10 for the outcome variable in the third column of Table 3. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table A12: Combined heterogeneity analysis: Incorrect highlights in required paragraphs

	(1)	(2)	(3)	(4)
Treatment	0.092	0.165	0.189	0.447
	(0.257)	(0.350)	(0.378)	(0.480)
Treat*Minority women	0.703	0.639	0.629	0.648
	(0.454)	(0.455)	(0.507)	(0.507)
Treat*Minority men	-0.203	-0.275	-0.421	-0.391
	(0.545)	(0.542)	(0.570)	(0.574)
Treat*Spelling confidence	—	0.121	0.110	0.093
		(0.393)	(0.402)	(0.401)
Treat*Grammar confidence	—	-0.671*	-0.664*	-0.643
		(0.391)	(0.394)	(0.396)
Treat*Science confidence	—	0.817	0.883	0.852
		(0.535)	(0.564)	(0.569)
Treat*High prevalence	—	—	-0.024	-0.027
			(0.443)	(0.443)
Treat*Past discrimination	—	—	0.045	0.061
			(0.403)	(0.404)
Treat*Income>80k	—	—	—	0.346
				(0.451)
Treat*Employed	—	—	—	-0.238
				(0.419)
Treat*College degree	—	—	—	-0.395
				(0.445)
N	1295	1295	1236	1236
p-value for null hypothesis that...				
No race/gender het.:	0.237	0.263	0.241	0.243
No gender het. (minorities):	0.155	0.145	0.103	0.109
No race het. (women):	0.122	0.161	0.215	0.202

Note: This table tests for heterogeneity along all possible explanatory dimensions; sequentially adding potential explanatory variables. The table is analogous to Appendix Table A10 for the outcome variable in the fourth column of Table 3. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table A13: Accounting for multiple hypothesis testing for gender and racial heterogeneity results

	Number of paragraphs	Correct highlights	Incorrect highlights	Res. wage (same mgr)
	(1)	(2)	(3)	(4)
White women (N=738):				
Non-blind manager effect	-0.572	-0.104	0.061	0.029
Unadjusted *p*-value	0.045	0.695	0.798	0.016
Adjusted *p*-value	0.341	0.975	0.798	0.161
Racial minority women (N=266):				
Non-blind manager effect	-0.874	0.785	0.753	-0.010
Unadjusted *p*-value	0.090	0.109	0.044	0.612
Adjusted *p*-value	0.525	0.491	0.357	0.975
Racial minority men (N=291):				
Non-blind manager effect	0.901	1.353	-0.132	0.022
Unadjusted *p*-value	0.096	0.003	0.776	0.324
Adjusted *p*-value	0.500	0.036	0.951	0.852

Note: This table reports treatment effects of being in the non-blind manager arm in the promotion experiment on the primary outcomes with heterogeneity by race × gender, accounting for multiple hypothesis testing with heterogeneity using the procedure developed in List et al. (2019). The outcomes are the number of paragraphs proofread, the number of correct highlights and incorrect highlights, and the reservation wage for future work when workers would be evaluated by the same manager. The multiple hypothesis testing adjustment accounts for 12 hypotheses (four outcomes and three groups) and uses 3,000 bootstrap iterations. The sample is the same as the corresponding columns in Tables 2, 3, and 4. The specification estimates the effects separately for each group, rather than jointly with the control variables constant across groups due to the construction of the List et al. algorithm. This means the estimates are slightly, but not meaningfully, different, from those in the main tables. The same control variables are used as in the main tables.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

Table A14: Effects on self-reported interest in future work

	Work for employer	Tasks w/ assignment	Easy paragraphs	Hard paragraphs	Prefer hard job
	(1)	(2)	(3)	(4)	(5)
Non-blind manager, see avatars	-0.058**	-0.040	-0.031	-0.048	-0.021
	(0.029)	(0.030)	(0.030)	(0.030)	(0.028)
N	1286	1286	1286	1286	1295
Control mean	0.649	0.557	0.582	0.529	0.684

Note: This table estimates the effect of being in the non-blind manager arm on self-reported interest in future work, relative to the demographic-blind manager arm in the promotion experiment. In columns 1-4, the outcomes are indicators for strongly agreeing with statements that they would like to do more tasks (on Prolific) for this employer, more tasks with this type of job assignment, more tasks of the same difficulty level, and tasks that are the harder job that they didn't do in the experiment. The outcome in column 5 is an indicator for probably or definitely preferring the harder job to the easier job. Specifications are the same as in Table 2. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.4 (Effects on Future Labor Supply)]

Table A15: Effects on future labor supply, gender and racial heterogeneity

	RW (Same)	WTP cutoff	Coop. RW	WTP new mgr	Shared zero	Amt shared
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Race and gender heterogeneity						
Non-blind manager*White women	0.033**	0.012**	0.024	0.050*	0.038	-0.245
	(0.014)	(0.006)	(0.019)	(0.027)	(0.039)	(0.316)
Non-blind manager*Minority women	-0.007	0.010	-0.020	0.114**	0.012	-0.619
	(0.021)	(0.011)	(0.030)	(0.044)	(0.062)	(0.448)
Non-blind manager*Minority men	0.026	0.013	-0.012	-0.019	0.109*	-0.076
	(0.023)	(0.012)	(0.032)	(0.046)	(0.058)	(0.495)
p-values:						
White=minority, women	*0.107*	*0.902*	*0.200*	*0.206*	*0.707*	*0.477*
Men=women, minority	*0.283*	*0.862*	*0.844*	*0.038*	*0.241*	*0.410*
N	1247	1235	1256	957	1289	1289
Panel B: Gender heterogeneity						
Non-blind manager*Women	0.022*	0.011**	0.012	0.066***	0.031	-0.345
	(0.012)	(0.005)	(0.017)	(0.023)	(0.034)	(0.270)
Non-blind manager*Minority men	0.026	0.013	-0.012	-0.019	0.109*	-0.076
	(0.023)	(0.012)	(0.032)	(0.046)	(0.058)	(0.495)
p-values:						
Men=women	*0.876*	*0.894*	*0.489*	*0.097*	*0.231*	*0.622*
N	1247	1235	1256	957	1289	1289
Panel C: Racial heterogeneity						
Non-blind manager*White women	0.033**	0.012**	0.024	0.050*	0.039	-0.244
	(0.014)	(0.006)	(0.019)	(0.027)	(0.039)	(0.316)
Non-blind manager*Minority	0.010	0.012	-0.016	0.051	0.062	-0.338
	(0.016)	(0.008)	(0.022)	(0.032)	(0.043)	(0.340)
p-values:						
White=minority	*0.251*	*0.996*	*0.155*	*0.984*	*0.670*	*0.831*
N	1247	1235	1256	957	1289	1289
Control Mean, white women	0.253	-0.009	0.325	0.205	0.529	2.815
Control Mean, minority women	0.280	-0.019	0.349	0.218	0.550	2.693
Control Mean, minority men	0.280	-0.026	0.366	0.264	0.517	2.924

Note: This table estimates the effect of being in the non-blind manager arm on future labor supply relative to the demographic-blind manager arm in the promotion experiment, separately for white women, racial minority women, and racial minority men. The outcomes and sample are the same as in Panel A of Table 4, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A8. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

[Click to return to Section 3.4 (Effects on Future Labor Supply)]

Table A16: Balance table by whether see two or three white men previously promoted

	Non-blind manager			Blind manager			Algorithm + avatars			Algorithm, no avatars		
	<3/3	3/3	*p*	<3/3	3/3	*p*	<3/3	3/3	*p*	<3/3	3/3	*p*
Male	0.22	0.22	*0.994*	0.26	0.17	*0.007*	0.22	0.27	*0.341*	0.27	0.23	*0.372*
Race:												
Asian	0.13	0.12	*0.464*	0.16	0.09	*0.013*	0.20	0.12	*0.099*	0.15	0.16	*0.891*
Black	0.15	0.13	*0.298*	0.15	0.14	*0.913*	0.13	0.18	*0.298*	0.20	0.18	*0.776*
White, Hispanic	0.15	0.15	*0.882*	0.16	0.17	*0.629*	0.16	0.11	*0.283*	0.16	0.16	*0.912*
White, Non-Hispanic	0.56	0.61	*0.262*	0.54	0.59	*0.154*	0.51	0.59	*0.200*	0.48	0.50	*0.837*
Married	0.50	0.51	*0.851*	0.55	0.53	*0.686*	0.53	0.55	*0.838*	0.60	0.52	*0.160*
Kids	0.49	0.39	*0.011*	0.45	0.43	*0.582*	0.40	0.48	*0.166*	0.45	0.40	*0.359*
Education:												
Less than high school	0.01	0.01	*0.814*	0.01	0.01	*0.767*	0.02	0.00	*0.154*	0.01	0.02	*0.692*
High school graduate	0.16	0.13	*0.268*	0.14	0.15	*0.763*	0.11	0.14	*0.383*	0.11	0.11	*0.861*
Some college but no degree	0.26	0.27	*0.818*	0.28	0.29	*0.828*	0.24	0.32	*0.169*	0.23	0.26	*0.499*
2 year college degree	0.09	0.11	*0.351*	0.10	0.09	*0.766*	0.08	0.07	*0.751*	0.10	0.08	*0.671*
4 year college degree	0.35	0.35	*0.827*	0.36	0.38	*0.734*	0.43	0.38	*0.430*	0.42	0.37	*0.357*
Professional or Masters degree	0.12	0.12	*0.878*	0.10	0.08	*0.380*	0.10	0.08	*0.660*	0.13	0.15	*0.677*
Doctorate	0.02	0.01	*0.330*	0.01	0.01	*0.972*	0.02	0.00	*0.218*	0.01	0.02	*0.692*
Income:												
Less than $20,000	0.14	0.16	*0.362*	0.12	0.13	*0.804*	0.15	0.15	*0.830*	0.09	0.22	*0.001*
$20,000-$40,000	0.21	0.20	*0.754*	0.22	0.28	*0.083*	0.24	0.18	*0.187*	0.25	0.17	*0.088*
$40,000-$60,000	0.20	0.19	*0.692*	0.18	0.17	*0.681*	0.23	0.19	*0.408*	0.18	0.11	*0.107*
$60,000-$80,000	0.14	0.16	*0.427*	0.15	0.14	*0.708*	0.13	0.16	*0.492*	0.16	0.12	*0.287*
$80,000-$100,000	0.09	0.12	*0.221*	0.13	0.08	*0.078*	0.08	0.10	*0.469*	0.09	0.10	*0.724*
$100,000-$120,000	0.10	0.05	*0.038*	0.07	0.08	*0.813*	0.04	0.07	*0.267*	0.07	0.09	*0.416*
$120,000-$140,000	0.03	0.02	*0.256*	0.04	0.04	*0.812*	0.04	0.03	*0.659*	0.05	0.05	*0.940*
$140,000-$160,000	0.03	0.05	*0.272*	0.03	0.04	*0.424*	0.02	0.04	*0.316*	0.05	0.03	*0.362*
More than $160,000	0.06	0.04	*0.401*	0.06	0.05	*0.539*	0.07	0.07	*0.879*	0.06	0.10	*0.216*
Age:												
18-24	0.15	0.19	*0.193*	0.17	0.17	*0.789*	0.20	0.13	*0.155*	0.19	0.16	*0.558*
25-34	0.28	0.35	*0.069*	0.31	0.28	*0.371*	0.29	0.40	*0.056*	0.31	0.34	*0.585*
35-44	0.25	0.18	*0.051*	0.22	0.23	*0.888*	0.26	0.15	*0.036*	0.19	0.20	*0.772*
45-54	0.15	0.13	*0.537*	0.16	0.14	*0.584*	0.13	0.10	*0.516*	0.13	0.15	*0.544*
55-64	0.11	0.10	*0.551*	0.09	0.13	*0.090*	0.06	0.15	*0.006*	0.14	0.11	*0.391*
65 or older	0.06	0.05	*0.587*	0.05	0.04	*0.698*	0.06	0.05	*0.849*	0.05	0.04	*0.721*
Employment:												
Currently employed outside Prolific	0.64	0.65	*0.738*	0.67	0.65	*0.598*	0.68	0.77	*0.095*	0.72	0.63	*0.116*
Ever employed outside Prolific	0.93	0.97	*0.033*	0.96	0.96	*0.935*	0.97	0.99	*0.278*	0.95	0.95	*0.815*
Satisfied with current employer	0.77	0.80	*0.454*	0.77	0.78	*0.751*	0.74	0.80	*0.337*	0.76	0.73	*0.646*
Satisfied with most recent employer	0.59	0.68	*0.189*	0.64	0.66	*0.799*	0.61	0.71	*0.391*	0.71	0.62	*0.369*
N	356	294		392	253		193	97		194	131	

Note: This table shows means of worker demographic characteristics, all measured in the baseline survey, by treatment group in the promotion experiment, separately for workers who were randomly paired with managers who previously promoted three white men or not (in the manager arms) or randomly paired with a group of twenty-four historical-sample workers jointly evaluated by the algorithm among whom three white men were previously promoted (in the algorithm arm). The sample includes racial-minority men and women and white women who would not have been promoted by any of the three procedures. *p*-values test the null of equality between workers who do or do not see three white men in each treatment group, and are calculated with standard errors robust to heteroskedasticity.

[Click to return to Section 4 (Effects of Minority-Group Representation)]

Table A17: Effects on other reasons for non-promotion in manager sample of the promotion experiment

	All Workers	Women	Minority Men	White Men
Panel A: Free response, first answer is more school				
Non-blind manager,	-0.106***	-0.116***	-0.152***	0.130
see avatars	(0.024)	(0.028)	(0.058)	(0.107)
Control mean	0.414	0.410	0.428	0.408
Panel B: Free response, first answer is higher quiz score				
Non-blind manager,	-0.058**	-0.078**	0.010	-0.059
see avatars	(0.027)	(0.031)	(0.066)	(0.114)
Control mean	0.484	0.488	0.476	0.469
Panel C: Free response, first answer is it was random				
Non-blind manager,	0.001	0.007	0.002	-0.098
see avatars	(0.011)	(0.013)	(0.021)	(0.066)
Control mean	0.045	0.046	0.028	0.082
N	1389	1004	291	94

Note: This table shows the effects of being in the non-blind manager arm on the other reasons that workers gave for what needed to be different about their profile to be assigned to the harder job (the question underlying the main measure of perceived discrimination), relative to the demographic-blind manager arm in the promotion experiment. The sample is workers in either manager arm who would not have been promoted under any of the three procedures. In Panel A the outcome is an indicator for thinking they needed more education to be promoted, in Panel B it is an indicator for thinking they needed higher quiz scores, and in Panel C it is an indicator for thinking it was random and there was nothing they could do. Specifications are otherwise the same as Figure 6. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.2 (Effects of Perceived Discrimination)]

Table A18: Effects on secondary measures of perceptions of discrimination in the algorithm sample of the promotion experiment

	All Workers	Women	Minority Men	White Men
Panel A: Free response, primary reason is discrimination				
See avatars	0.085***	0.103***	0.053	–
	(0.019)	(0.025)	(0.042)	
Control mean	0.003	0.004	0.000	0.000
Panel B: Yes, would be promoted if different race or gender				
See avatars	0.164***	0.200***	0.145*	0.002
	(0.030)	(0.037)	(0.078)	(0.065)
Control mean	0.054	0.045	0.096	0.022
Panel C: Complaint about job assignment mentions discrimination				
See avatars	0.039***	0.041***	0.043	–
	(0.012)	(0.015)	(0.031)	
Control mean	0.000	0.000	0.000	0.000
*Panel D: Members of my race * gender group underrepresented among promoted*				
See avatars	0.109***	0.119**	0.103	0.025
	(0.040)	(0.049)	(0.104)	(0.108)
Control mean	0.294	0.285	0.446	0.065
Panel E: White men are overrepresented among promoted				
See avatars	0.127***	0.137***	0.234**	-0.243
	(0.044)	(0.053)	(0.110)	(0.197)
Control mean	0.472	0.479	0.494	0.391
N	691	464	151	76

Note: This table shows the effects of seeing that previously-promoted workers were mostly white men in the algorithm sample of the promotion experiment on secondary measures of perceived discrimination. The secondary measures are the same as in Appendix Table A6. The sample is workers in the algorithm arm who would not have been promoted under any of the three procedures. Specifications are otherwise the same as in Figure 7. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.2 (Effects of Perceived Discrimination)]

Table A19: Effects on psychological outcomes in the promotion experiment

	Total good affect (SD)	Self-efficacy index (SD)	Very satisfied with work
	(1)	(2)	(3)
Non-blind manager, see avatars	0.014	-0.007	-0.028
	(0.037)	(0.042)	(0.030)
N	1295	1288	1295
Control mean	0.013	-0.016	0.549

Note: This table estimates the effect of being in the non-blind manager arm on psychological well-being, relative to the demographic-blind manager arm in the promotion experiment. The first outcome is an index of overall affect (emotional state) and the second is the average of two self-efficacy indices, one for skills related to the proofreading job and one over the proofreading jobs themselves. Both are measured in standard deviation units. The third is an indicator for reporting being very satisfied with the job in the experimental survey. The specifications and sample are the same as in Table 2. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

Table A20: Effects on psychological outcomes, gender and racial heterogeneity

	Total good affect (SD)	Self-efficacy index (SD)	Very satisfied with work
	(1)	(2)	(3)
Panel A: Race and gender heterogeneity			
Non-blind manager*White women	0.002	0.012	-0.008
	(0.048)	(0.052)	(0.039)
Non-blind manager*Minority women	-0.059	-0.050	-0.052
	(0.079)	(0.094)	(0.062)
Non-blind manager*Minority men	0.111	-0.015	-0.057
	(0.072)	(0.084)	(0.060)
p-values:			
White=minority, women	*0.499*	*0.551*	*0.536*
Men=women, minority	*0.110*	*0.780*	*0.954*
N	1295	1288	1295
Panel B: Gender heterogeneity			
Non-blind manager*Women	-0.014	-0.004	-0.020
	(0.042)	(0.047)	(0.034)
Non-blind manager*Minority men	0.111	-0.015	-0.057
	(0.071)	(0.084)	(0.060)
p-values:			
Men=women	*0.123*	*0.908*	*0.579*
N	1295	1288	1295
Panel C: Racial heterogeneity			
Non-blind manager*White women	0.002	0.012	-0.008
	(0.048)	(0.052)	(0.039)
Non-blind manager*Minority	0.029	-0.032	-0.055
	(0.054)	(0.064)	(0.044)
p-values:			
White=minority	*0.702*	*0.577*	*0.407*
N	1295	1288	1295
Control Mean, white women	0.038	-0.043	0.572
Control Mean, minority women	-0.063	-0.079	0.471
Control Mean, minority men	0.026	0.111	0.566

Note: This table estimates the effect of being in the non-blind manager arm on future labor supply relative to the demographic-blind manager arm in the promotion experiment, separately for white women, racial minority women, and racial minority men. The outcomes and sample are the same as in Panel A of Appendix Table A19, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A8. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.
[Click to return to Section 3.3 (Effects on Retention and Performance)]

B Robustness analyses

B.1 Alternative inference methods

B.1.1 Randomization inference

Randomization inference is a standard method used to determine the statistical significance of an estimate without making assumptions about its asymptotic distribution (Fisher, 1935; Rosenbaum, 2002). Specifically, I re-assign treatment status using the same procedure as in the experiments 500 times, indexed by $k \in (1,500)$, and for each iteration k, estimate the parameter of interest $\hat{\beta}^k$ using the re-assigned "treatment" with the same specification as the main analysis. The distribution of $\hat{\beta}^k$'s provides the empirical distribution of the treatment effect parameter under the null hypothesis that $\beta = 0$. The two-sided randomization inference p-value is the fraction of absolute values of $\hat{\beta}^k$'s that are greater than the absolute value of the observed $\hat{\beta}^{true}$ estimated using the real treatment assignment. Almost all of the results and conclusions of the paper are unaffected.

Promotion experiment: Panel A Appendix Figure B1 plots the p-values of the effects of assignment by a non-blind manager in the promotion experiment on all of the outcomes of interest. Of the outcomes for which the asymptotic standard errors suggest a statistically significant effect at a less than 5 percent level, all of the randomization-inference p-values remain below 0.05; in most cases, the randomization inference-based p-values are smaller than the asymptotic ones. The p-value on whether workers share with their manager increases from 0.11 to 0.19 when using randomization inference. Panel B is analogous for the effects of learning that the demographic-blind algorithm previously promoted mostly white men; the effects on completing any optional paragraphs and performance have randomization inference-based p-values below 0.07. Finally, the effects of increasing minority-group representation retain p-values less than 0.03 (Appendix Figure B2).

Hiring experiment: Appendix Figure B3 plots randomization-inference p-values for the *differences* in rates of perceived discrimination between the arms of the hiring experiment. The effects of blinding decision-makers to demographics and the effect of using an algorithm instead of a manager when the decision-maker is non-blind remain significant at the 5 percent level; the effect of using an algorithm instead of a manager when the decision-maker is demographic-blind remains significant at the 10 percent level.

B.1.2 Multiple hypothesis testing

All of the figures referenced above also plot corresponding p-values when they are corrected for multiple hypothesis testing; most results maintain significance at a 10 percent level. Specifically, I control the family-wise error rate using the Romano-Wolf (2005; 2016) correction with 500 bootstrap iterations.

Promotion experiment: In Appendix Figure B1 the "families" are the groups in bold: retention, effort and performance in the first six paragraphs, future labor supply (including cooperation and sharing with managers), and beliefs about future promotion (in the manager arm only). The effects of being in the non-blind manager arm relative to the demographic-blind manager arm on future labor supply maintain p-values between 0.08 and 0.12; the effects on beliefs maintain p-values below 0.02. The positive effect on number of paragraphs proofread does not maintain significance, however, this result is driven by racial minority men and that result does maintain significance when corrected for multiple hypothesis testing across both the number of main outcomes and the number of heterogeneous groups tested (Appendix Table A13).

The main results that are not robust to using MHT-corrected inference are the effects of learning that the demographic-blind algorithm previously promoted mostly white men on performance; these p-values

increase to more than 0.2 when they are adjusted. Appendix Figure B2 does not include MHT-corrected p-values for the effects of seeing that three white men were previously promoted as there are only two estimates in the "family." A simple Bonferroni-style correction that multiplied the asymptotic p-values by two would imply p-values that are still less than 0.02.

Hiring experiment: The effects of blinding managers to demographics and the effect of using an algorithm instead of a manager have adjusted p-values around 0.07; the effect of blinding an algorithm remains significant at the 1 percent level (Appendix Figure B3).

B.2 Attrition and attention

Differential attrition and attention also cannot explain any of the results.

Attrition: Appendix Table B1, columns (1) and (2) shows that there is no differential attrition in the promotion experiment between the demographic-blind and non-blind manager arms or between workers in the algorithm sample that do or do not see demographics (Panels A and B). Panels C and D show that there is also no differential attrition between workers who saw three white men or two white men and someone else in the non-blind manager arm or in the demographic-blind algorithm arm when workers saw previously-promoted workers' demographics, respectively. Appendix Figure B4 shows that there is similarly no differential attrition in the hiring experiment.

Attention: The same tables and figures show that there is no difference in the rate at which workers failed the attention check in the final part of the experimental survey, except between workers who saw three white men versus two white men and someone else in the algorithm + avatars arm of the promotion experiment (Appendix Table B1, Panel D, column 3). Note that workers who failed the attention check are included in the main sample but the results are robust to dropping them, discussed next. Section 2.4 shows that worker attention to and comprehension of the experiment was high overall.

B.3 Specification choices

I test the sensitivity of the main estimates to the choice of control variables and sample definition by estimating the main results while iterating over all possible decision-points in Appendix Figures B5-B11.

B.3.1 Sample definitions

The main results restrict the sample to workers who were not selected by any of the procedures, as pre-registered. This eliminates concerns that different procedures may have selected different types of workers, generating differences between the treatment arms in the non-promoted or non-hired sample other than those caused by different perceptions of the reason they were not selected. The main results in the promotion experiment also drop the small sample of white men who were recruited as a comparison group to ensure that the treatment triggered perceived race and gender discrimination; this was not pre-registered. The specification charts iterate over these and other sample definitions: further restricting to workers who passed the "attention check" in the final survey section, including the white men in the promotion experiment results, or expanding to the full sample of workers who completed the experimental surveys.

The effects of perceived manager discrimination on behavior are generally unchanged, but effects of perceived algorithmic discrimination are more sensitive. Specifically, the effects of learning that the demographic-blind algorithm previously promoted mostly white men on earnings and performance are less negative and lose significance when restricting to those that passed the attention check; the positive effects on retention disappear for that sample as well (Appendix Figure B9). The effects of anti-bias hiring procedures on perceived discrimination and effects of minority-group representation on perceived discrimination are unchanged.

B.3.2 Included controls

The main specification controls for demographic information (age, education, income, employment, race, gender), baseline quiz scores, quiz-score group fixed effects, and the educational experience and quiz-score quintiles of the previously-promoted workers each worker saw.

A second specification (in the manager arms only) replaces the quiz-score group fixed effects with evaluation group fixed effects (groups within which workers were jointly assigned to be evaluated by the same counterfactual managers) but leaves the rest of the main specification controls unchanged. I pre-registered using evaluation-group fixed effects rather than quiz-score group fixed effects. I primarily present results using quiz-score group fixed effects because this allows for the same exact specification to be used for the manager and algorithm samples (the non-blind/blind manager randomization was within evaluation groups, whereas the with/without avatars algorithm randomization was across evaluation groups). As seen here, this does not matter for the results.

An alternate set of controls was selected by a double-post-lasso procedure, following Chernozhukov et al. (2018) as pre-registered. Specifically, I run a linear lasso model separately for each outcome and treatment-group indicators including all possible control variables as independent variables (about 350, each standardized to have mean zero and standard deviation one).[1] These models are run separately for those evaluated by a manager and those evaluated by an algorithm and separately for each sample definition in the previous subsection. For each outcome $\times$ sample, the lasso-selected controls are the union of controls selected by lasso when predicting the outcome in that sample and when predicting treatment assignment in that sample. The baseline survey was shortened for the hiring experiment to minimize costs so I observe fewer covariates, and only carry out this exercise for the promotion experiment.

A fourth set of controls removes the controls for the education and performance of the previously-promoted workers, since these were not pre-registered. As seen here, this does not matter for the results. They are included in the main specification because there are significant differences in these variables across treatment arms. Finally, the figures also show estimates from regressions that include no control variables.

The estimated treatment effects of evaluation by a non-blind manager are very stable regardless of the choice of control variables, consistent with effective randomization and the lack of observable differences between the treatment arms. This is true for all four sets of results (the effects of perceived manager discrimination or perceived algorithmic discrimination on worker behavior or the effects of anti-bias hiring policies and minority-group representation on rates of perceived discrimination). Unsurprisingly, adding evaluation-group fixed effects tends to improve the precision of the estimated coefficient in the manager arms of the promotion experiment, and removing all controls does the opposite for all results.

B.4 Demand effects

A common critique of survey experiments is the potential for differential experimenter demand effects: if participants perceive that they are in an experiment, figure out the treatments, and guess the researcher's hypothesis, differences between treatment arms could be induced by a desire to help (or harm) the researcher's "agenda" (de Quidt et al., 2019). That said, recent evidence suggests that quantitative social scientists running survey experiments (and their readers) may be unnecessarily or overly concerned with

[1] All controls in the main specification, plus: indicators for reporting having experienced gender or racial discrimination in past job search, promotion, termination, or daily work activities, second-order beliefs about those probabilities for each of eight race $\times$ gender groups, job satisfaction with their current or most recent employer, whether they've ever been unemployed and looking for work, whether they've experienced an unemployment spell longer than six weeks, scales of racial and gender identity centrality, and family status. The set of possible controls also includes 250 indicators for and interactions of the elements of the profiles that non-blind managers had access to when making decisions about who to assign to higher-status job: 11 self-identified race and ethnicity categories, gender, having a college degree or more than a college degree, quiz-score quintiles, five skin tones, eight hair colors, and 24 hairstyles, and interactions of race/ethnicity $\times$ gender $\times$ quintiles $\times$ degree status as shown on the profiles.

bias from experimenter demand effects, relative to an older literature of lab experiments in psychology in which the phenomena was originally documented.[2]

Unsurprisingly, Appendix Table B2 (promotion experiment) and Appendix Figure B12 (hiring experiment) show that regardless of when they were asked, workers in the arms that made the possibility of discrimination salient were much more likely to say that they thought the study was about discrimination.[3] In the promotion experiment 0 percent and 3 percent of the control (no salient demographics) and treatment (salient demographics) groups, respectively, know that the study is about perceived discrimination when asked after the proofreading task, and 0-3 and 4-13 percent say the same when asked after the reservation wage, willingness to pay, and dictator game elicitations. A randomly-selected half of the sample was asked at each of these times. All workers are asked again at the end of the study. Now, 35-40 percent of the control groups think the study was about discrimination, and being in a treatment arm where the possibility of discrimination is salient increased this by 8pp (se=4.5pp) and 11pp (se=3pp) when evaluated by the algorithm or a manager, respectively. This is likely due to the final questions in the survey that ask about perceptions of discrimination most explicitly. Among those evaluated by a non-blind manager, workers were 7-10pp more likely to report that they thought the study was about discrimination when they saw three white men than when they saw two white men and someone else were previously promoted; workers in the demographic-blind algorithm arm that learned previously-promoted workers' demographics were 5-10pp more likely to do so (before the explicit perceived discrimination elicitations).

In the hiring experiment, workers are asked what they thought was the topic of the study first after the reservation wage elicitations and again at the end of the survey. At both times, workers are more likely to think that the study was about discrimination in the arms in which there were positive rates of perceived discrimination (Appendix Figure B12). At the first elicitation, 0-3 percent of workers in the control groups and 16-24 percent of workers in the groups where the possibility of discrimination was salient think the study was about discrimination. At the second elicitation, this was 8-18 percent and 30-40 percent, respectively.

Even when workers differentially think that the study was about discrimination in the treated groups, there still may not be confounding demand effects if workers do not alter their behavior based on this knowledge. I estimate conservative bounds on all of the main results that account for the possibility of experimenter demand effects by replacing the value of each outcome variable with its value plus or minus 0.2 times the standard deviation of that variable in the demographic-blind manager arm if the worker guesses that the topic of the study was discrimination (for the outcomes measuring worker behavior) and, more conservatively, I replace the perceived discrimination outcome with 0 or 1 if the worker guesses that the topic of the study was discrimination.[4] Appendix Figure B13 present the results for the effect of being in the non-blind manager arm relative to the demographic-blind manager arm on worker behavior with these bounds. By subtracting (adding) 0.2sd to the outcome, I obtain an estimate for the effect of treatment if

[2] Mummolo and Peterson (2019) show that giving participants information about experimenter hypotheses does not affect participant behavior or estimated treatment effects, even when participants are financially incentivized to conform to those hypotheses and participants are more knowledgable about the study purpose than participants in the "no information" condition. de Quidt et al. (2018) show that telling participants they will "do the researcher a favor" by choosing a higher or lower action than they otherwise would generates substantial bias in estimated treatment effects, and saying that the researcher "expects participants shown these instructions to [choose a higher or lower action] than they normally would" generates some, but much less, bias.

[3] These indicators were generated via coding workers' answers to the open-ended question, *"What do you think this study is about? It's fine if you're uncertain, please still write something down."* The independent variable in these regressions is an indicator for workers' answers being about discrimination. In the promotion experiment, these were coded by external coders. Due to budget and time constraints, in the hiring experiment, I generated the coded variables from the text responses.

[4] de Quidt et al. (2018) show that a strong message telling workers to work harder or less hard changes outcomes by about 0.1sd in the incentivized effort task they test, though they are not powered to detect this small of an effect. They do not test effects on multiple price list elicitations like the ones I use to measure reservation wages, which are less cleanly incentivized. So, as a conservative bound, I use 0.2 as my multiplier on the standard deviation. I use an even more conservative bound on the perceived discrimination outcome which is not incentivized at all.

participants who know the study topic have a higher (lower) value of the outcome than they would have if they did not know the topic. There are three upper and three lower bounds; these use the first, second, and third elicitation of beliefs about the study purpose to determine whose outcomes to replace (recall, the first and second elicitations were each asked of half of the sample). The treatment-effect point estimates are remarkably similar and my conclusions are unchanged, though some of the lower bounds are no longer statistically significant. Appendix Figure B14 shows that the effects of perceived algorithmic discrimination on performance are also robust to this adjustment.

Appendix Figure B15 shows the analogous bounds for the effects of minority-group representation on perceived discrimination rates, now replacing the perceived discrimination measure with 0 or 1 if a worker believed the study was about discrimination. The lower bound is about half the size of the main point estimate and the upper bound is similar to the main point estimate. Thus, a more conservative interpretation would be that increasing minority-group representation from 0 to 1 of 3 cuts rates of discrimination by about 20-25 percent rather than about 40-50 percent.

In the hiring experiment, I estimate similar bounds on the differences between rates of perceived discrimination in each of the treatment arms, replacing the perceived discrimination indicator with a 0 or 1 if a worker thinks the study was about discrimination. I use knowledge of the study topic measured after the reservation wage elicitation (which was asked of all workers). The lower bound on the effect of blinding managers and algorithms to demographics are is again about half the size of the main estimate and only marginally significant. The effects of using an algorithm rather than a manager are less affected but also smaller (Appendix Figure B16).

B.5 Multiple price list elicitations

The four measures of future labor supply were elicited using multiple price lists (MPLs).[5] These were: reservation wages for another round of evaluation and the same work when (1) workers would be evaluated by the same manager as in the experiment and promoted or not, and (2) when baseline quiz scores would alone determine promotion; reservation wages to do a collaborative task with their manager; and willingness to pay to be able to choose a different manager to work with in the collaborative task. The exact implementation of each MPL is available in the *Online Supplement.*

Following the literature, these were incentivized by telling workers that I would randomly choose one row of the MPL and implement the worker's choice in that row for randomly-selected workers. That said, workers could report low reservation wages and just not take the job later. This was uncommon and does not seem to matter for the results: only 10-15 percent of workers answered "I would want the job at this wage" in all rows for each of the three reservation wages, there is no treatment effect on whether workers answer in this way, and dropping these workers does not change the estimated treatment effects on reservation wages.

Some workers may not have understood or paid attention to the instructions for the MPL questions. In the main results, I follow the literature and drop any worker who does not respond to the MPL questions in a sensible way: these are workers who display multiple switching points throughout the list or those who switch in the wrong direction (i.e, indicate downward-sloping labor supply or upward-sloping demand for choosing their own manager). Appendix Table B3 compares the workers who are dropped for this reason with other workers in the promotion experiment. These workers are less attentive and perform worse on the proofreading task than those who answer the MPL questions sensibly, but have similar education (Panel A). Consistently, the treatment effects on all other outcome variables are driven by the workers who *do* answer the MPL questions sensibly (Panel B), with the exception of whether and how much workers share with their manager.

[5]Budget constraints prohibited me offering all workers a second round of work to measure effects on *observed* future labor supply.

B.6 Deviations from the Pre-Analysis Plan

Exploiting random variation in the fraction of previously-promoted workers who are white men. I could not predict which workers managers or the algorithm would select out of twenty-four randomly-grouped historical-sample workers. I did not anticipate much variation in what fraction of previously-promoted workers were white men. However, there was substantial variation, with 40 percent of workers in the experimental sample seeing three white men being previously promoted and the rest seeing two white men and someone else. Workers were randomly paired with managers or groups of workers jointly evaluated by the algorithm, and indeed seeing that all previously-promoted workers are white men is not correlated with any observable worker characteristics (Appendix Table A16). The documented differences in rates of perceived discrimination between workers who saw three white men previously promoted and those who saw two white men and someone else have important implications and thus are included prominently in the paper despite not being pre-registered. There were also differences across arms in what fraction of workers saw three white men; Figure 11 controls for this difference and allows for comparisons across evaluation procedures conditional on whether workers saw two or three white men previously promoted.

Given these results, a second related deviation from the pre-analysis plan was to focus the analysis in the hiring experiment on the subsample of workers who see only three white men previously promoted. All of the main exhibits are duplicated with the full sample in the appendix and all secondary analysis is available upon request in the full sample.

Change in sample composition. The demographic make-up of the sample also deviates from the pre-analysis plan. I managed to over-sample racial minorities relative to the population but not to the extent originally planned. I also exhausted a larger fraction of the white men interested in my study in recruiting the historical sample than anticipated, and so I recruited fewer white men to the experimental sample than intended. I recruited additional white women to the sample until the pre-registered overall sample size was reached. Another deviation is that I drop the small number of white men who participate in the promotion experiment from the majority of the analysis after showing that the treatment does not generate perceived discrimination for white men (as expected), but results are very similar when they are added to those analyses.

Change from pre-registered sample size. First, I recruited 120 fewer workers to the screening survey for the promotion experiment than pre-registered. This was mainly because of difficulty recruiting the 75 managers, since they were required to be white men (the pool of which had been largely exhausted in recruiting the historical sample). Second, in order to recruit as many racial minority participants as I did, I had to raise the base wage paid for the screening survey for the last part of the sample. Stopping recruitment to the screening survey early helped make up this unexpected expense in the budget. All of this occurred before any workers were randomly assigned to treatment groups or took the experimental survey.

Second, in both the promotion and hiring experiments, one group of workers (one evaluation group of 120 (360) in the promotion (hiring) experiment) was not offered the experimental survey after completing the screening survey because I exhausted all possible managers that could be brought back to evaluate them and none did. Since they had no data on counterfactual decision-making by that manager, they would not have been able to be in the main analysis sample. Since treatment was assigned within evaluation groups, this had no implications for the analysis. Again, this decision was made before any workers were assigned a treatment status. It also helped to make up the above-mentioned slight budgetary shortfall.

Added controls for previously-selected workers' education and baseline performance. In the control and treatment groups, workers saw the education and baseline performance of the workers that their manager or the algorithm had selected in the historical sample. Comparing the treated groups, which also saw previously selected workers' race, gender, and avatar, to the control group nets out the effect of knowing that one was not selected and that previous workers with similar quiz scores as you were previously selected. However, the different evaluation procedures selected workers in the historical sample with systematically

different education and baseline performance. To account for these differences—for example, to make sure that the effects are driven by perceived racial and gender discrimination rather than feeling mis-evaluated because one did not have a college degree—I added controls for the education and baseline performance of the three previously-selected workers to the baseline specification. As discussed in Section B.3 however, this does not affect the main conclusions. Estimates without these controls are in Appendix Figures B5-B11.

A mistake in the PAP's description of control variables. The PAP described the main specification as including ν_i, fixed effects for the groups of workers jointly evaluated by one manager or the algorithm. This is not possible; treatment does not vary within these groups. The correct fixed effect is γ_i as described in Section 3.1: workers' baseline quiz-score level (taking on four values), which is the group within which treatment was randomly assigned.

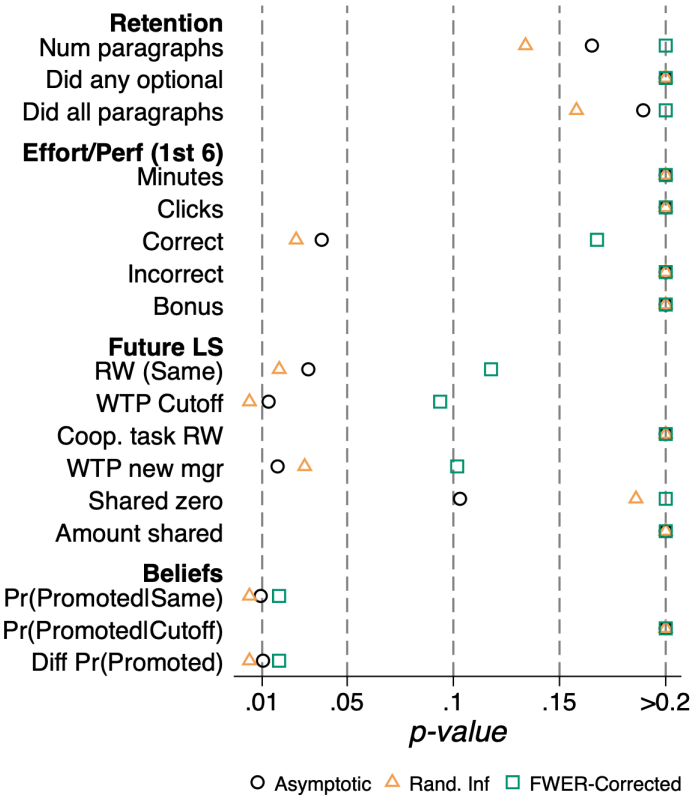
De-emphasis of IV specification. Section 3.1 discusses the challenges associated with interpreting the IV specification, despite it being pre-registered. The results are available in the *Online Supplement.*

Implicit measures of perceived discrimination. The paper focuses on one primary measure of perceived discrimination and shows results for the other pre-registered measures of explicit perceived discrimination, but does not show results for the pre-registered measures of implicit perceived discrimination. These are differences in the number of stars (quiz-score quintiles) that workers report thinking they would have needed to be promoted compared to members of different demographic groups. The signs of the treatment effects on these variables are consistent with the explicit measures, but imprecise (available upon request).

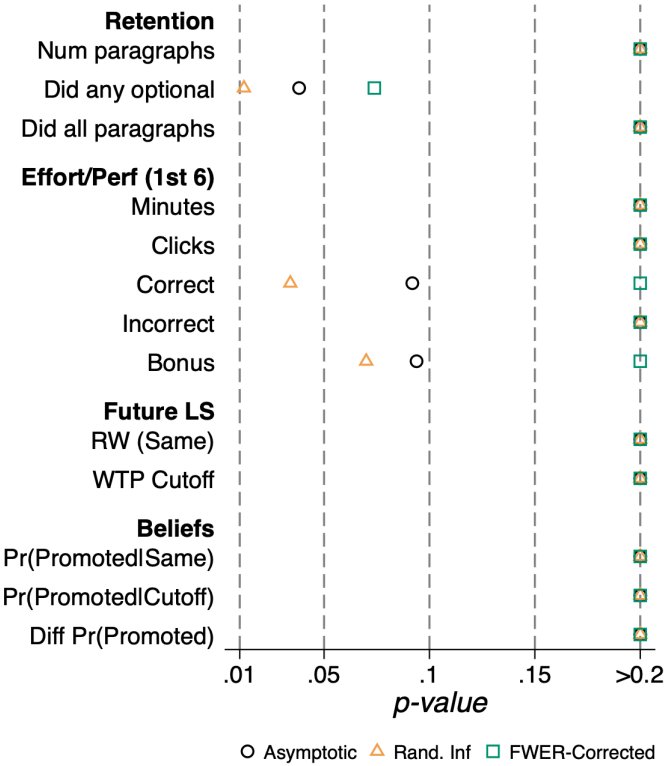
Standard error formulation. The pre-analysis plan outlined that I would cluster standard errors by the groups of 40 workers that were jointly evaluated by the same manager. However, random assignment to each manager was at the individual level so this was incorrect. If anything, clustering standard errors in this case makes them smaller, not larger.

Figure B1: Randomization inference and multiple hypothesis testing

(a) Effects of perceived manager discrimination

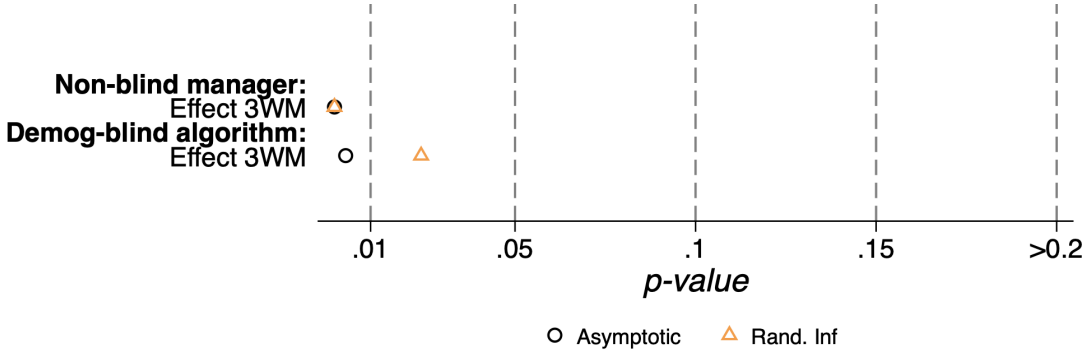


(b) Effects of perceived algorithmic discrimination



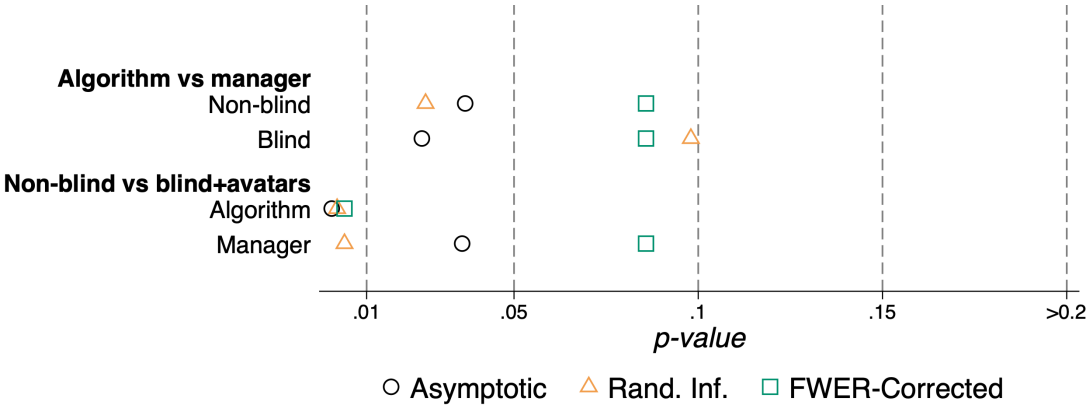
Note: These figures plots the *p*-values for the effects of being in the non-blind manager group relative to the demographic-blind manager group (on the left) and the effects of seeing that the demographic-blind algorithm previously promoted mostly white men (on the right) on the key outcomes of interest, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity), with randomization inference, and corrected for multiple hypothesis testing using the Romano-Wolf correction.

Figure B2: Randomization inference and multiple hypothesis testing (historical representation)



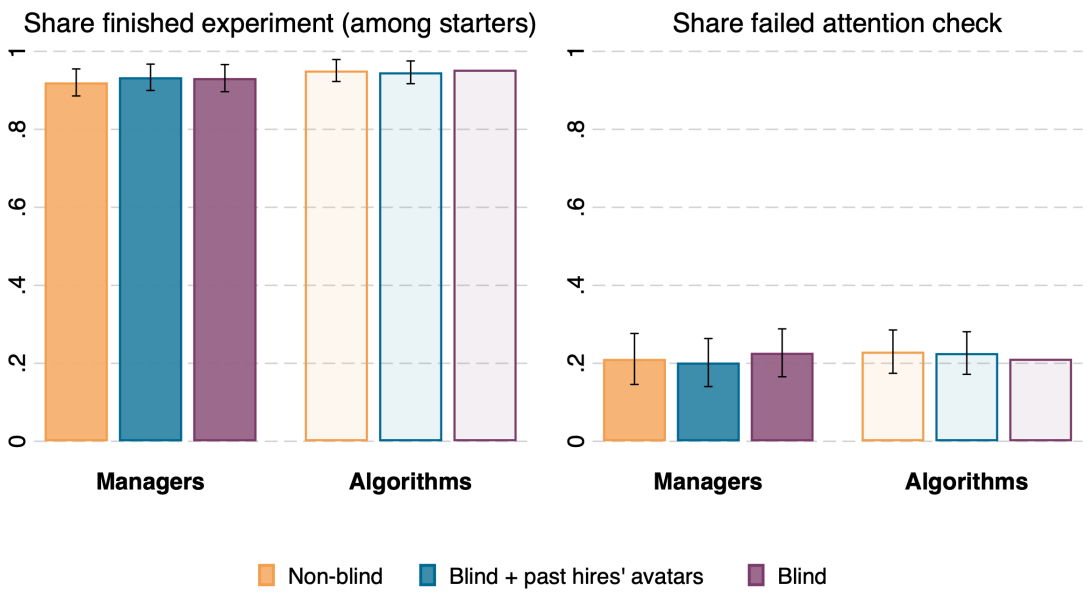
Note: This figure plots the *p*-values for the differences in perceived discrimination between treatment arms in Figure 11, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity) and with randomization inference.

Figure B3: Randomization inference and multiple hypothesis testing (hiring experiment)



Note: This figure plots the *p*-values for the differences in perceived discrimination between treatment arms in Figure 12, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity), with randomization inference, and corrected for multiple hypothesis testing using the Romano-Wolf correction.

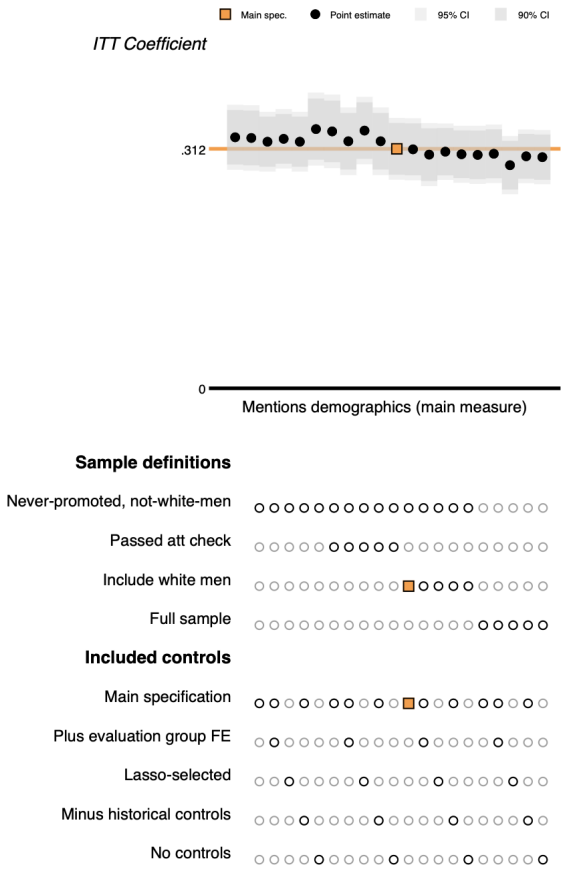
Figure B4: Attrition and attention in the hiring experiment



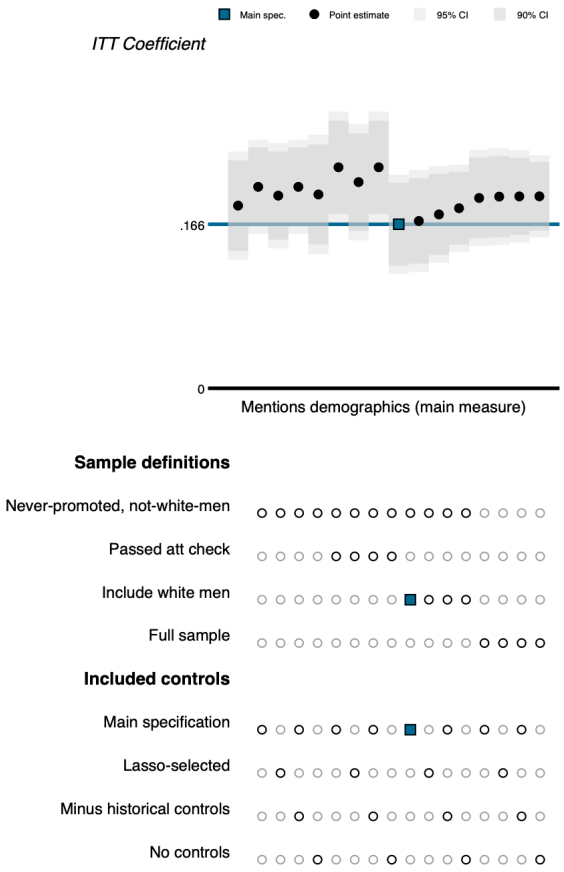
Note: This figure shows the fraction of workers in each arm of the hiring experiment who finished the experiment, among those who started it (on the left) and the share of the main sample that failed the attention check (on the right). The sample is restricted to workers who would not have been promoted under any hiring procedure, and workers who saw that all three previous hires were white men (a random subsample within each treatment arm—results are the same for the full sample). Shares are calculated via the same regressions as in Figure 12. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

Figure B5: Alternative specifications: Effect of treatment on perceived discrimination (promotion experiment)

(a) Effect of being in non-blind manager arm

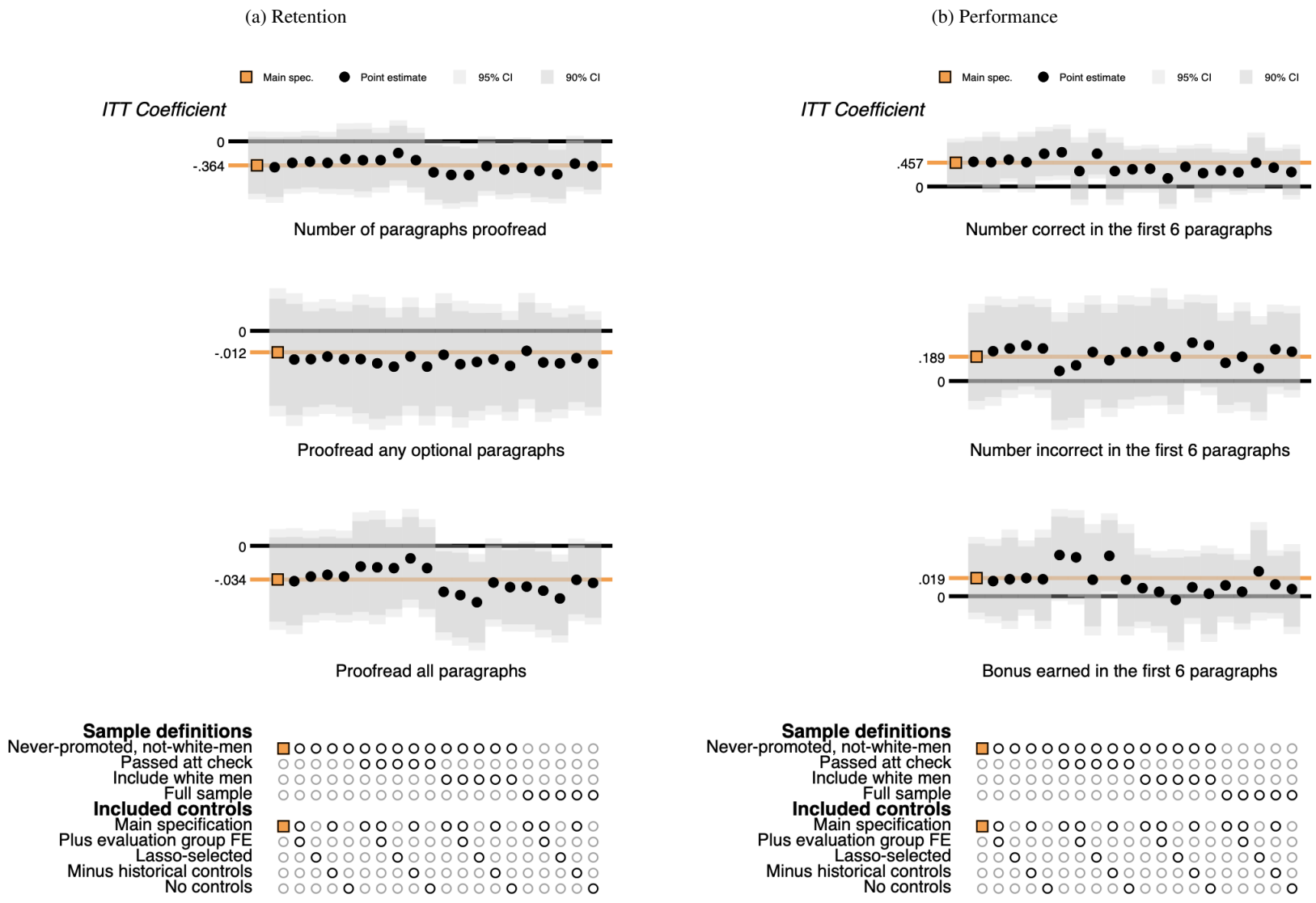


(b) Effect of learning algorithm previously promoted white men



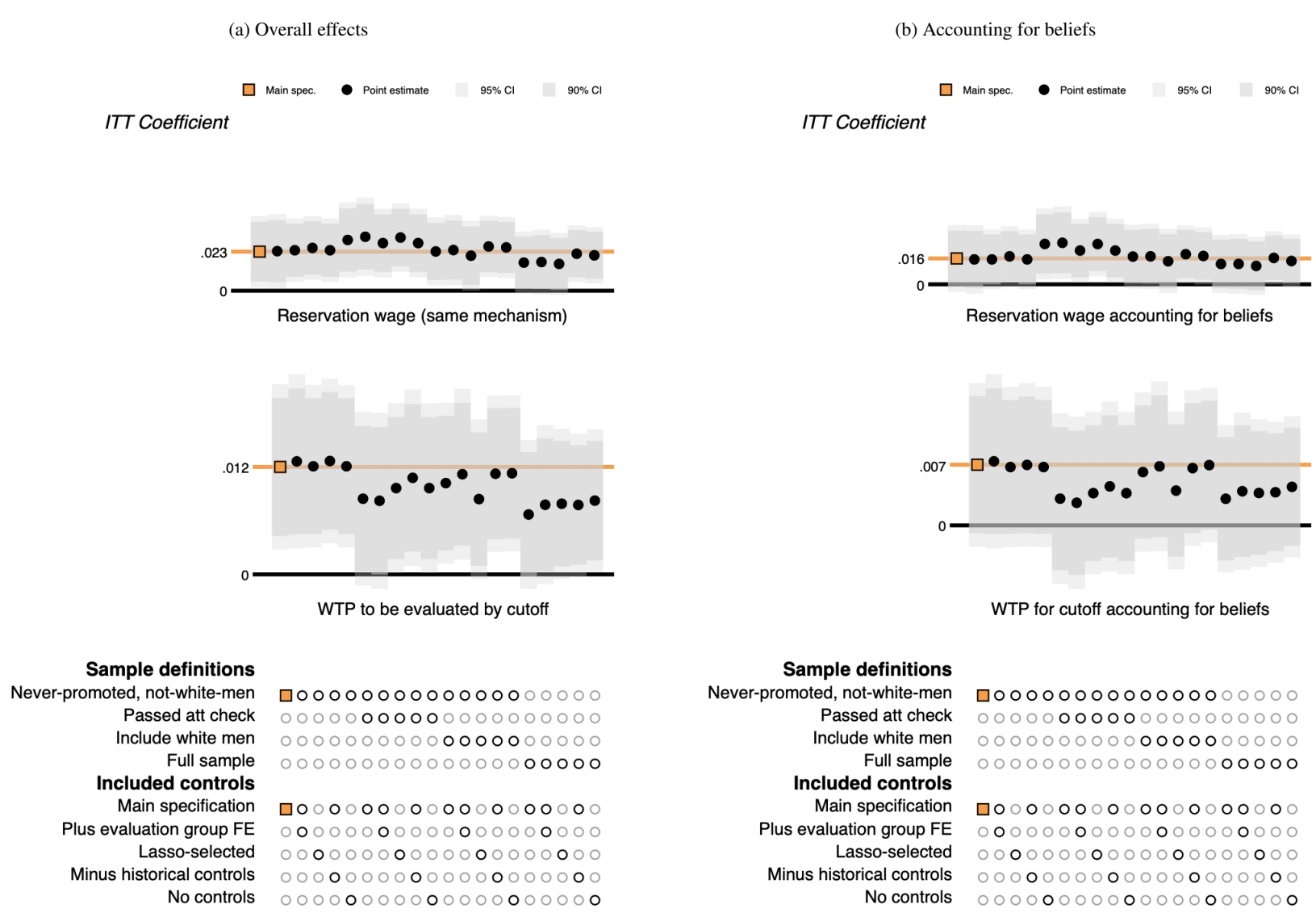
Note: This figure plots the effect of the two treatments in the promotion experiment on the main measure of perceived discrimination in the promotion experiment. The figure iterates over sample restrictions and choices of control variables. Starting with the main sample used in the analysis of the effects of perceived discrimination (workers who would not have been promoted under any procedure who are not white men), it then additionally restricts to workers who passed an attention check near the end of the survey, then expands to include never-promoted white men, and then the full sample (all workers who completed the experimental survey, including white men and workers who would have been promoted under any of the procedures). The controls begin with the main specification, then, in Panel A, add evaluation group fixed effects (the groups of 120 who were counterfactually evaluated by the same managers). Because the randomization within the algorithm arm was across rather than within evaluation groups, Panel B does not include a specification that adds evaluation group fixed effects. Then, I remove the controls for the composition of previous hires' education and performance, as they were not pre-registered. Next, I include double-post lasso-selected controls, and then drop all control variables. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure B6: Alternative specifications: Effect of perceived manager discrimination on retention and performance



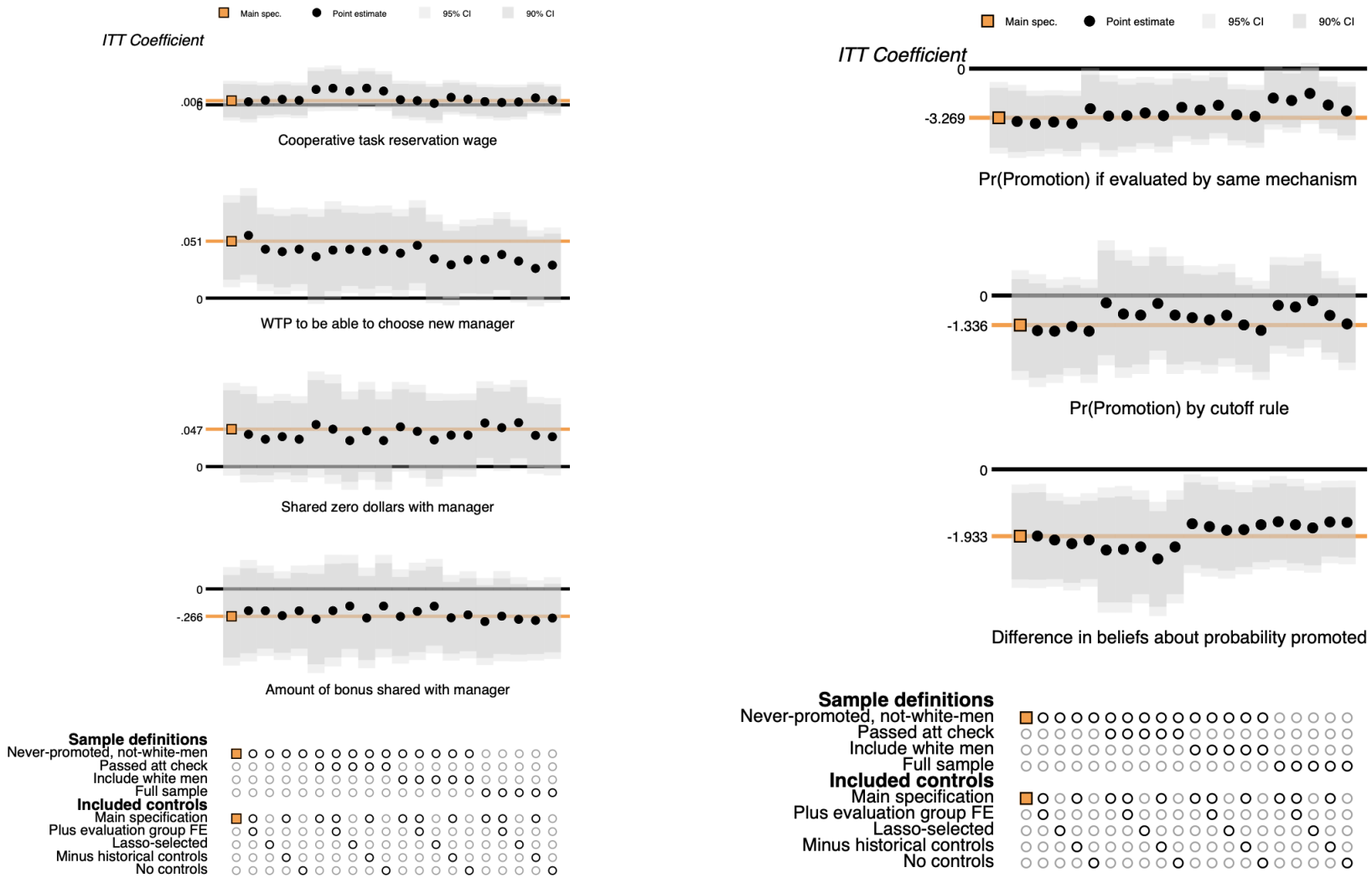
Note: This figure plots the treatment effect of being in the non-blind manager group on retention and performance in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Tables 2 and 3 (columns 3-5). Otherwise, the figure parallels Appendix Figure B5, Panel A.

Figure B7: Alternative specifications: Effect of being in non-blind manager arm on future labor supply



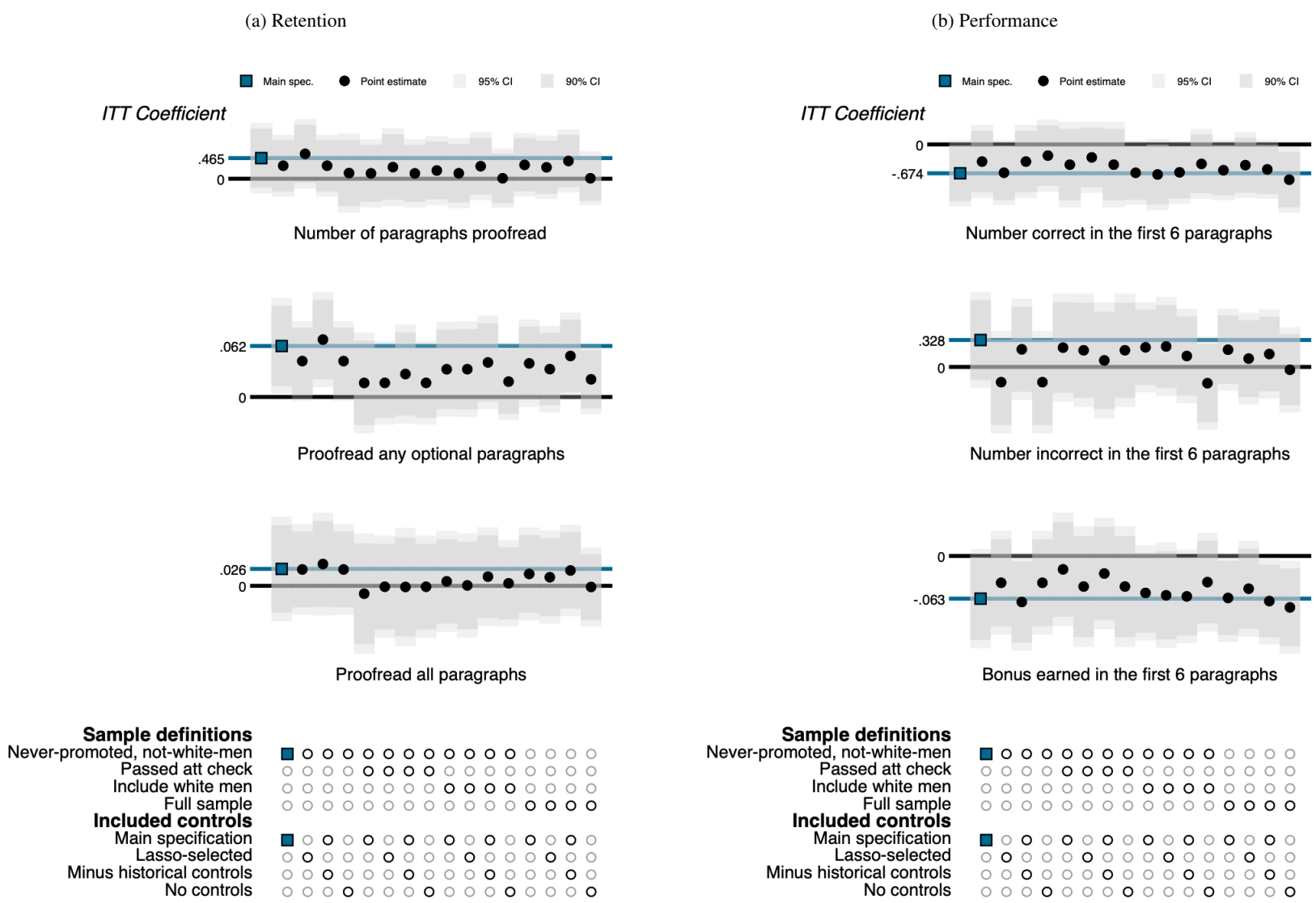
Note: This figure plots the treatment effect of being in the non-blind manager group on reservation wages in the promotion experiment, relative to the demographic-blind manager group. Panel A shows effects on the reservation wage and willingness to pay overall, while Panel B shows the effects after accounting for changes in beliefs about the likelihood of promotion. The outcomes and main specification are those in Figure 9. Otherwise, the figure is analogous to Appendix Figure B5, Panel A.

Figure B8: Alternative specifications: Effect of being in non-blind manager arm on cooperation, generosity, and beliefs



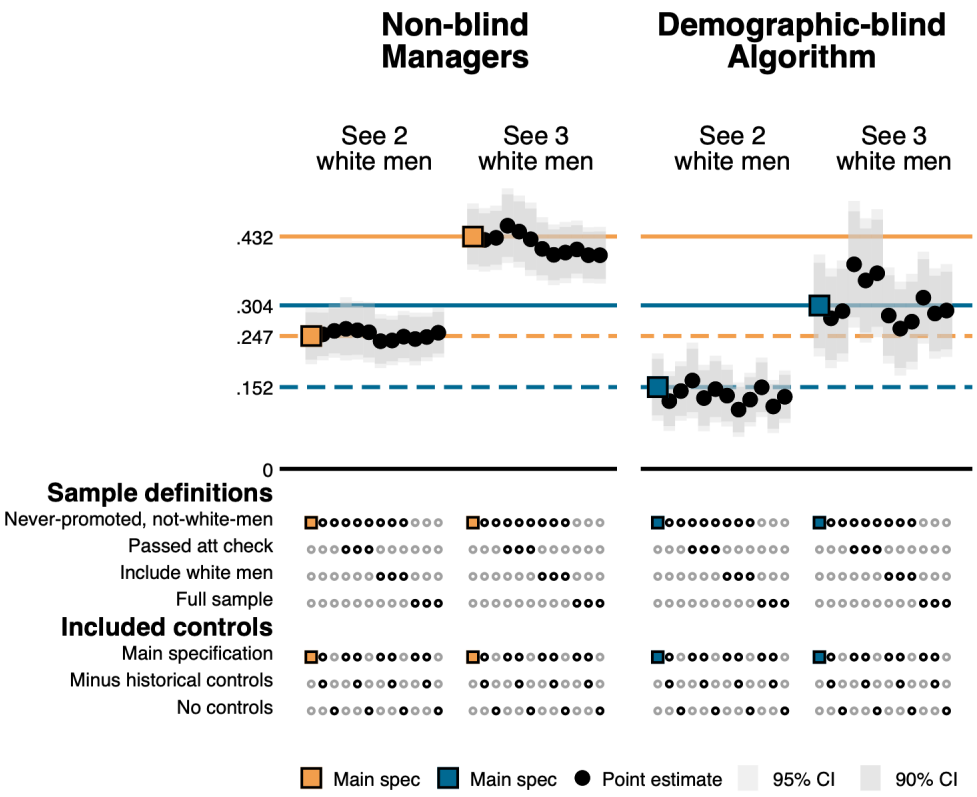
Note: This figure plots the treatment effect of being in the non-blind manager group on willingness to interact and share with managers in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Tables 5 and 6. Otherwise, the figure is analogous to Appendix Figure B5, Panel A.

Figure B9: Alternative specifications: Effect of perceived algorithmic discrimination on worker behavior



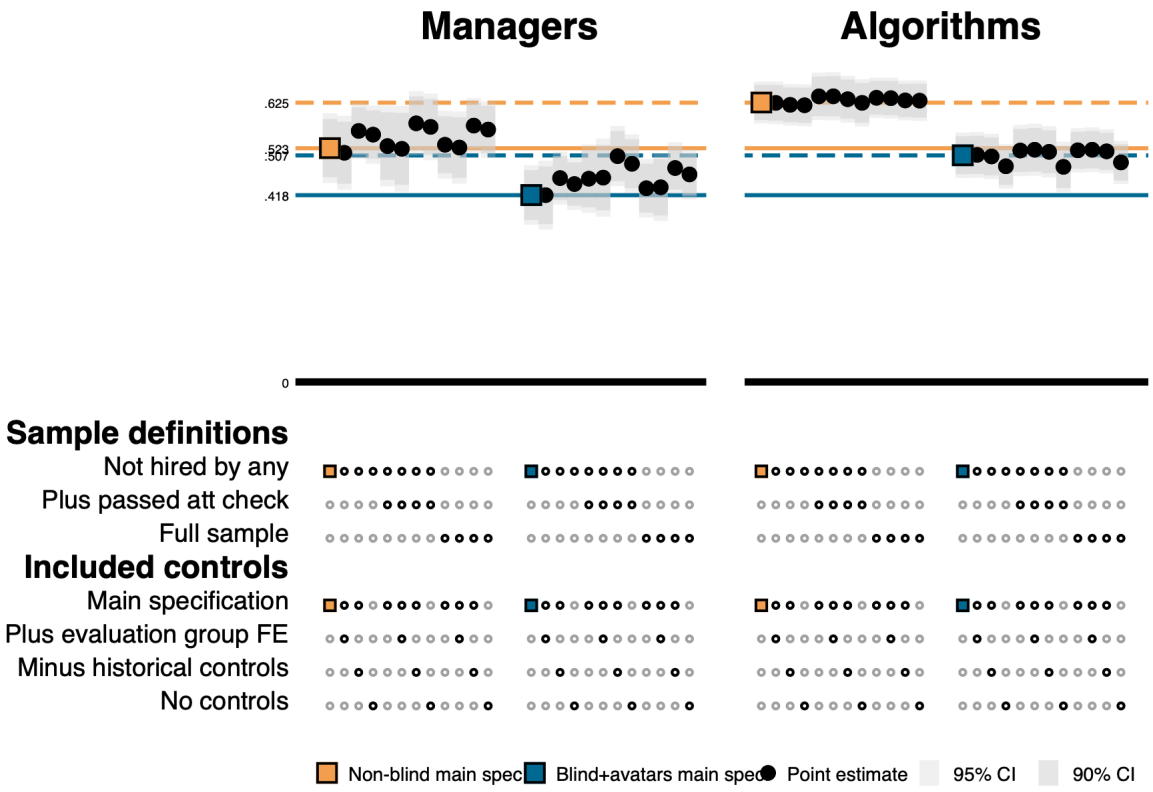
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on retention and performance in the avatar arm of the promotion experiment. The outcomes and main specification are those in Figure 10. Otherwise, the figure parallels Appendix Figure B5, Panel B.

Figure B10: Alternative specifications: Effects of minority-group representation



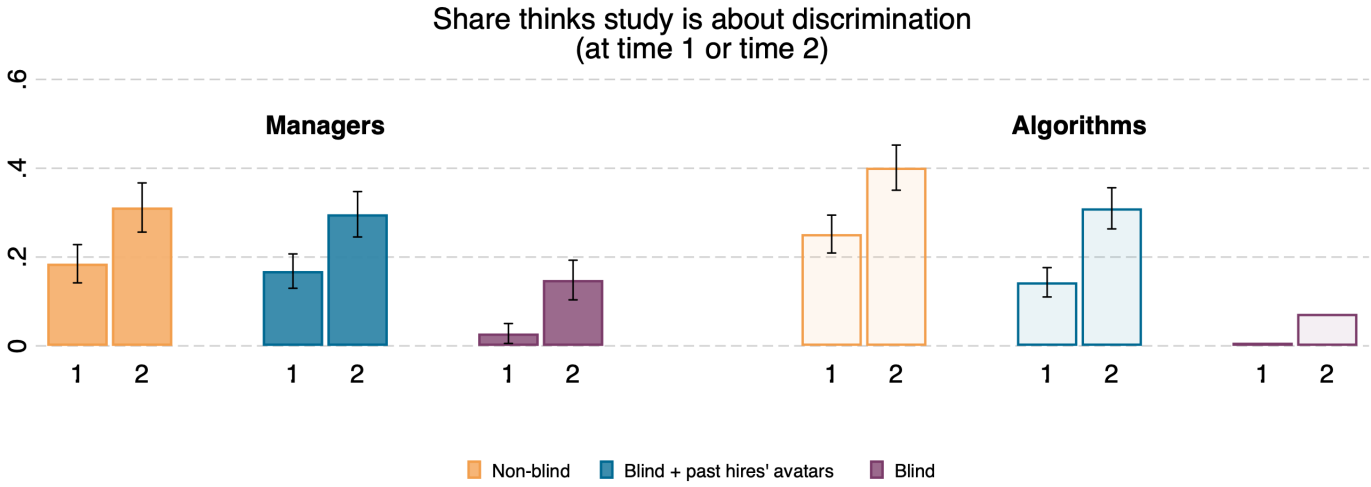
Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the promotion experiment. It parallels Figure 11, plotting only the share perceiving discrimination in the non-blind manager arm and the demographic-blind algorithm arm in which workers saw previously-promoted workers' demographics (though these rates are estimated using the same specification as Figure 11. The figure iterates over other sample restrictions and choices of control variables, starting with the main specification. Starting with the main sample definition (workers who would not have been hired under any procedure), it then additionally restricts to workers who passed an attention check near the end of the survey, then expands to the full sample. The controls begin with the main specification, then, I remove the controls for the composition of previous hires' education and performance, as they were not pre-registered. Next, I drop all control variables. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure B11: Alternative specifications: Hiring experiment (if previous hires are all white men)



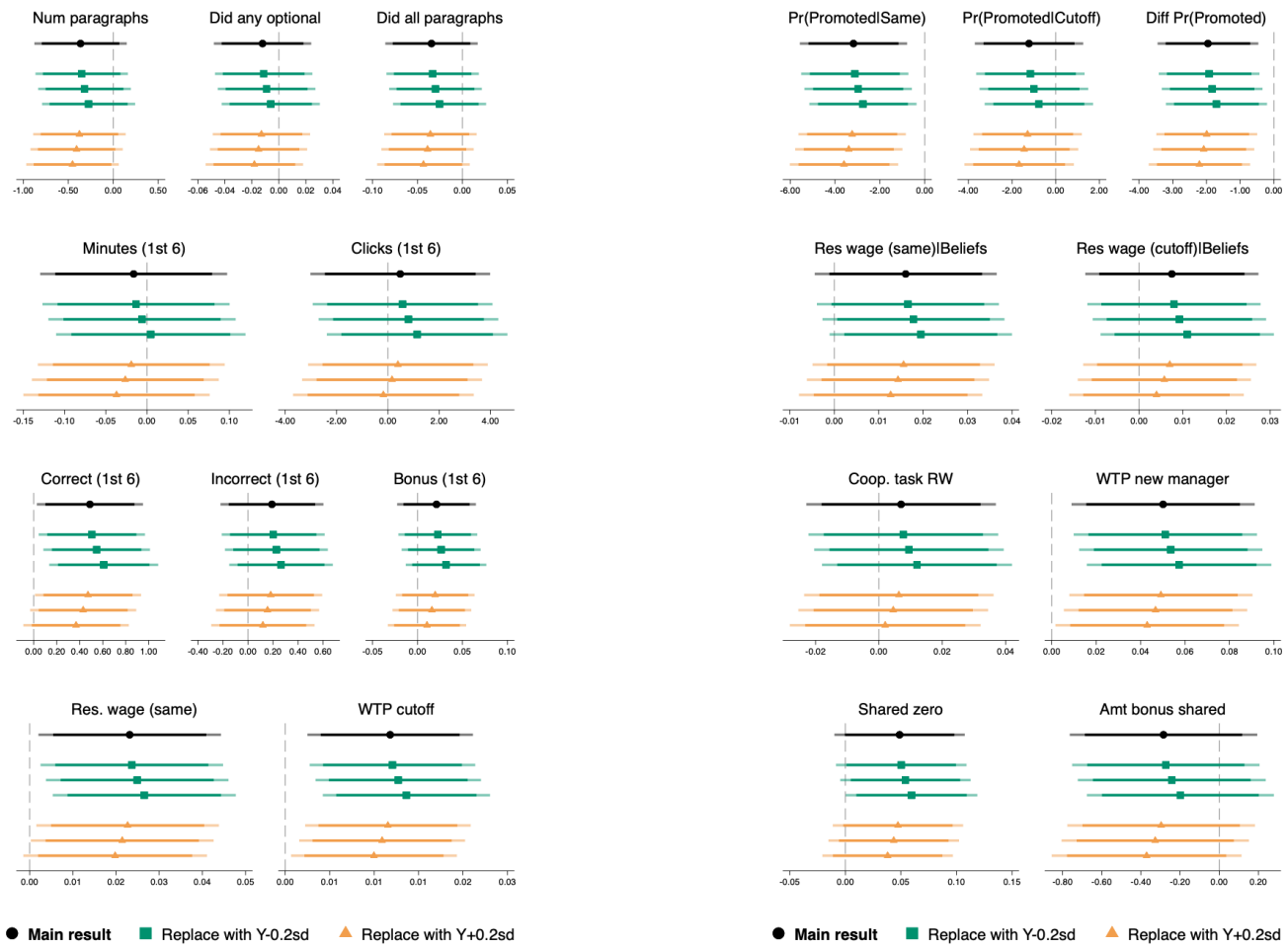
Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the hiring experiment. It parallels Figure 12, with the sample restricted to workers who saw only white men being previously hired. The figure iterates over other sample restrictions and choices of control variables, starting with the main specification. Starting with the main sample definition (workers who would not have been hired under any procedure), it then additionally restricts to workers who passed an attention check near the end of the survey, then expands to the full sample. The controls begin with the main specification, then add evaluation group fixed effects (the groups of 120 who were counterfactually evaluated by the same managers). Then, I remove the controls for the composition of previous hires' education and performance, as they were not pre-registered. Next, I drop all control variables. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure B12: Knowledge of study topic in the hiring experiment



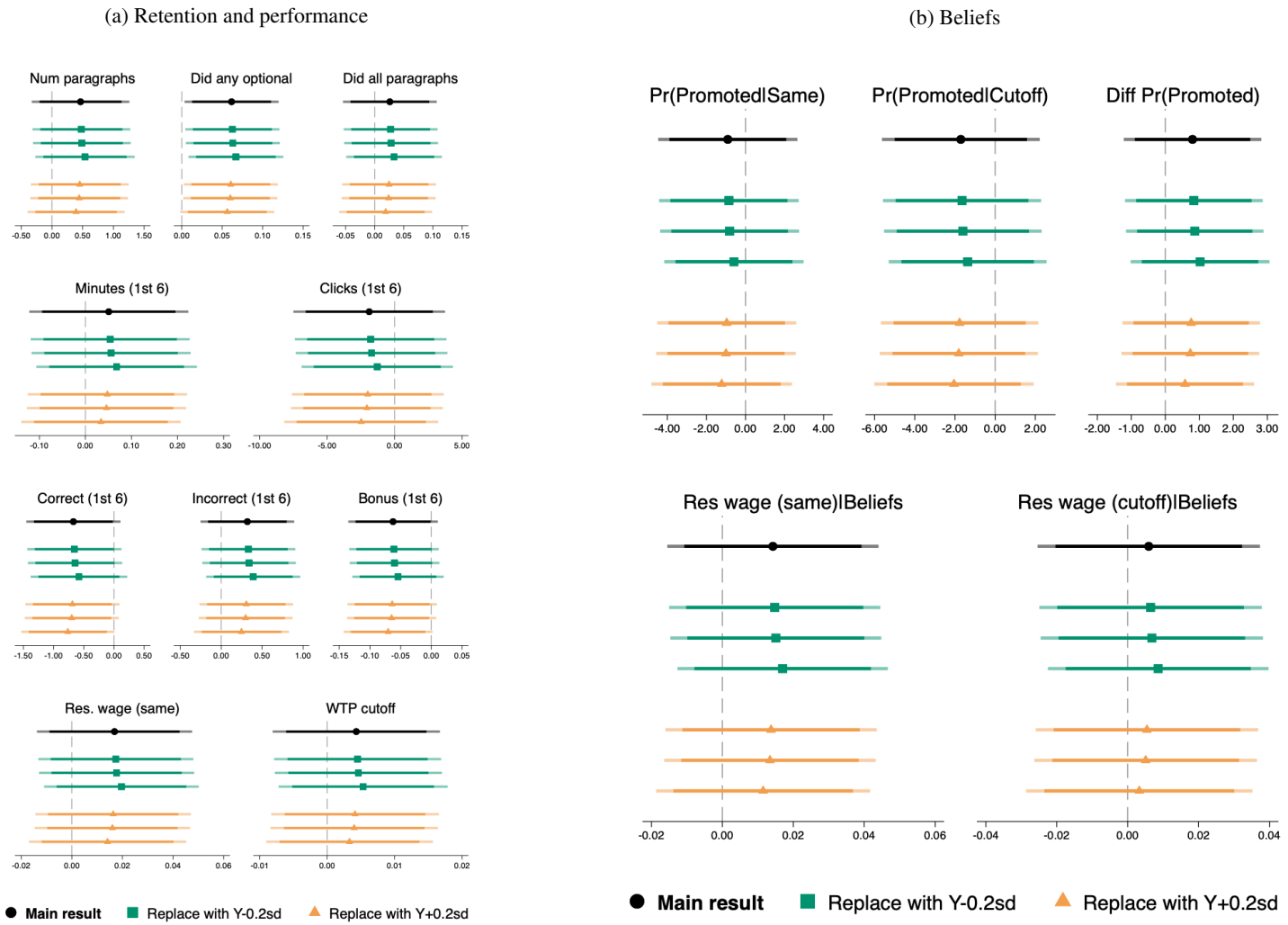
Note: This figure shows the fraction of workers in each arm of the hiring experiment who thought that the study was about discrimination the first time they were asked (after learning that they weren't hired and the reservation wage elicitations, before any survey questions) and the second (at the end of the survey). The sample is restricted to workers who would not have been promoted under any hiring procedure, and workers who saw that all three previous hires were white men (a random subsample within each treatment arm—results are the same for the full sample). Shares are calculated via the same regressions as Figure 12. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

Figure B13: Testing experimenter demand in the manager arms of the promotion experiment



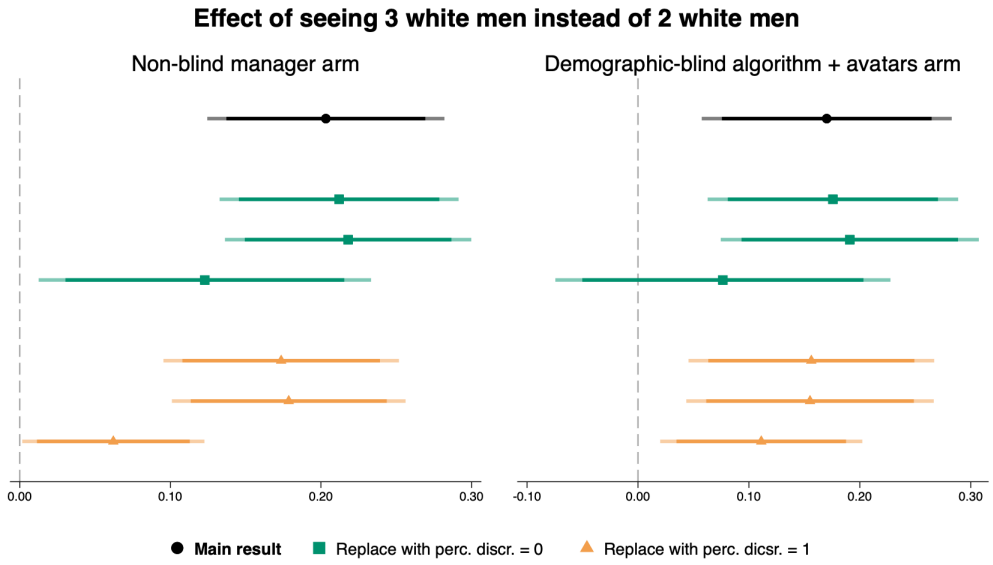
Note: This figure plots the treatment effects of being in the non-blind manager arm in the promotion experiment, relative to the demographic-blind manager arm, for all main outcomes with adjustments to account for possible experimenter demand effects. These bounds replace the outcome variable with its value plus or minus 0.2 standard deviations of that variable in the demographic-blind manager arm if workers thought that the study was about discrimination when they were asked after the proofreading task, willingness to pay elicitations, and at the end of the study (in the first, second, and third estimate for each outcome). The treatment effects estimated using these alternate outcome variables are plotted below the main estimate. This accounts for workers who knew the study topic potentially changing their answer to the question about perceived discrimination based on what they thought the researcher "wanted" to hear, differentially by treatment group. The specifications are otherwise identical to the main tables. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure B14: Testing experimenter demand in the algorithm arm of the promotion experiment



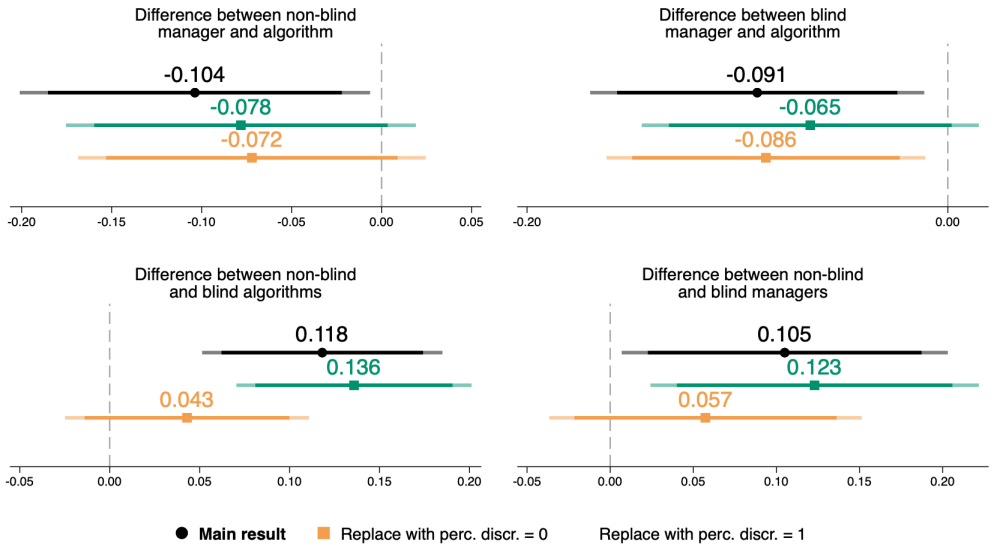
Note: This figure plots the treatment effects of learning the demographics of previously-promoted workers in the promotion experiment among workers evaluated by the demographic-blind algorithm with adjustments to account for possible experimenter demand effects, analogous to Appendix Figure B13.

Figure B15: Testing experimenter demand for effects of seeing versus two white men previously promoted



Note: This figure plots the effects of seeing that three white men were previously promoted from Figure 11, with adjustments to account for possible experimenter demand effects. These bounds replace the outcome variable (perceived discrimination) with a zero or one if workers thought that the study was about discrimination when they were asked after the proofreading task, willingness to pay elicitations, and at the end of the study (plotted first, second, and third, respectively). The specifications are otherwise identical to Figure 11. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure B16: Testing experimenter demand in the hiring experiment



Note: This figure plots the differences in perceived discrimination between treatment arms in Figure 12, with adjustments to account for possible experimenter demand effects. These bounds replace the outcome variable (perceived discrimination) with a zero or one if workers thought that the study was about discrimination when they were asked just after the elicitation of the outcome variable and the reservation wage elicitation. The specifications are otherwise identical to Figure 12. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Table B1: Attrition

	Starts experiment (baseline sample)	Finishes experiment (among starters)	Failed survey attention check
	(1)	(2)	(3)
Panel A: Manager arms of the promotion experiment			
Non-blind manager	-0.001	-0.004	-0.012
	(0.018)	(0.012)	(0.025)
N	1593	1354	1295
Control mean	0.848	0.951	0.248
Panel B: Algorithm arms of the promotion experiment			
See avatars	-0.029	0.030	-0.051
	(0.029)	(0.019)	(0.042)
N	763	649	615
Control mean	0.864	0.929	0.283
Panel C: Non-blind manager arm of the promotion experiment			
See three white men	-0.037	0.016	0.009
	(0.027)	(0.014)	(0.039)
N	793	676	650
Mean, see two white men	0.870	0.952	0.228
Panel D: Algorithm+avatars arm of the promotion experiment			
See three white men	-0.012	-0.017	0.168***
	(0.047)	(0.027)	(0.064)
N	358	299	290
Mean, see two white men	0.839	0.975	0.207

Note: This table tests for differential attrition in the promotion experiment. The sample always drops white men to match the primary analysis of the main experiment. In column 1, the sample is all workers who took the baseline survey and would not have been promoted under any procedure and the outcome is an indicator for starting the experimental survey. In column 2, the sample is all workers who start the experimental survey and would not have been promoted under any procedure, and the outcome is an indicator for finishing the experimental survey. In column 3, the sample is the main analysis sample of workers who finished the experimental survey and would not have been promoted under any procedure, and the outcome is an indicator for failing the attention check question near the end of the survey. Controls are the same as in all main analysis. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.

Table B2: Knowledge of study topic

	After proofreading	After WTP elicitations	End of survey
	(1)	(2)	(3)
Panel A: Manager arms of the promotion experiment			
Non-blind manager	0.031***	0.099***	0.108***
	(0.012)	(0.022)	(0.030)
N	651	644	1295
Control mean	0.009	0.032	0.395
Panel B: Algorithm arms of the promotion experiment			
See avatars	0.034*	0.043**	0.080*
	(0.019)	(0.018)	(0.045)
N	316	299	615
Control mean	0.000	0.000	0.345
Panel C: Non-blind manager arm of the promotion experiment			
See three white men	0.073***	0.070	0.098**
	(0.026)	(0.049)	(0.048)
N	321	329	650
Mean, see two white men	0.006	0.128	0.455
Panel D: Algorithm+avatars arm of the promotion experiment			
See three white men	0.052	0.100**	-0.012
	(0.049)	(0.050)	(0.072)
N	144	146	290
Mean, see two white men	0.011	0.041	0.404

Note: This table tests for differential knowledge of the study topic in the promotion experiment. The sample always drops white men to match the primary analysis of the main experiment. In column 1, the outcome is an indicator for thinking that the study is about discrimination after the proofreading job (before the reservation wage elicitations and other survey questions) in the random half of the sample that is asked at that time; in column 2, it is a similar indicator but measured after the reservation wage elicitations, in the other half of the sample; in column 3, it is an indicator for thinking the study is about discrimination at the end of the study (asked of the whole sample). Controls are the same as in all main analysis. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 level is indicated by *, **, and ***, respectively.

Table B3: Comparing workers with sensical and non-sensical MPL elicitations

	Dropped from any MPL elictation	Not dropped	*p-value* (1)=(2)
	(1)	(2)	(3)
Panel A: Means in the blind manager arm			
Indicators of worse attention:			
Non-standard open response	0.075	0.036	*0.031*
Failed att check 1 (screener)	0.090	0.065	*0.258*
Failed att check 2 (screener)	0.116	0.063	*0.023*
Failed att check (experiment)	0.372	0.189	*0.000*
Education:			
Less than high school	0.020	0.005	*0.057*
High school graduate	0.121	0.149	*0.344*
Some college but no degree	0.266	0.293	*0.493*
2 year college degree	0.106	0.097	*0.734*
4 year college degree	0.407	0.349	*0.159*
Professional or Masters degree	0.075	0.099	*0.336*
Doctorate	0.005	0.009	*0.596*
Proofreading performance, effort:			
Correct highlights, required paragraphs	12.407	15.714	*0.000*
Incorrect highlights, required paragraphs	4.176	3.061	*0.000*
Time on required paragraphs	4.859	4.982	*0.133*
N (blind manager arm)	199	444	
Panel B: Treatment effects (SEs)			
Number of paragraphs	0.459 (0.483)	-0.705 (0.294)	*0.032*
Proofread more than required	0.040 (0.035)	-0.032 (0.020)	*0.058*
Finished all paragraphs	0.029 (0.048)	-0.062 (0.030)	*0.095*
Time on required paragraphs	0.065 (0.105)	-0.050 (0.065)	*0.328*
Bonus in required paragraphs	0.068 (0.042)	-0.002 (0.025)	*0.149*
Beliefs about promotion, same mgr	-4.453 (2.066)	-2.474 (1.407)	*0.405*
Beliefs about promotion, cutoff	-4.529 (2.098)	0.155 (1.485)	*0.056*
Pr(same manager) - Pr(cutoff)	0.076 (1.289)	-2.628 (0.874)	*0.069*
Shared zero with manager	0.159 (0.053)	0.000 (0.034)	*0.010*
Amount shared with manager	-0.751 (0.458)	-0.064 (0.274)	*0.183*
N (both manager arms)	377	916	

Note: This table compares those who are dropped from the sample when looking at outcomes elicited using multiple price lists and those who are not, in the manager sample of the promotion experiment. The sample always drops white men to match the primary analysis of the main experiment. Workers are dropped in the main analysis of the MPL-based outcomes if they display multiple switching points or switch in the wrong direction (e.g. display downward-sloping labor supply). 80 percent of those who are dropped only display non-sensical responses for their willingness to pay to choose their own managers (the elicitation of which switched "direction" from the three previous reservation wage elicitations). Panel A compares the means of measures of attention and skill/effort in the blind manager arm across the two groups. Panel B shows the estimated treatment effect of being in the non-blind manager arm in each sample, estimated in one regression by interacting the treatment indicator with an indicator for being dropped in the main analysis. The specification is otherwise the same as Table 2. In panel B, standard errors are robust to heteroskedasticity (in parentheses).