Perceived discrimination at work

Hannah Ruebeck*

MIT

March 22, 2024

Abstract

Beliefs about experiencing discrimination are widespread but understudied. In an online experiment (N≈5000), I randomly assign workers to be evaluated by promotion procedures with varied potential to discriminate and provide information about the procedure. Learning that managers knew workers' race and gender and previously promoted mostly white men increases perceived discrimination rates from 3-33%, lowers retention by 3-6%, and increases reservation wages by 9%. Reducing perceived discrimination is therefore important for equity and efficiency. However, increasingly-common anti-bias procedures—blinding managers to demographics or using unbiased algorithms—are unlikely to alone eliminate perceived discrimination when minority groups remain under-represented.

*Email: hruebeck@mit.edu. I am immeasurably grateful for the support of Esther Duflo, Parag Pathak, and Frank Schilbach, who have advised this project since its earliest days. I would also like to thank Abi Adams-Prassl, David Autor, Abhijit Banerjee, Lisa Ho, Sendhil Mullainathan, Lucy Page, Ashesh Rambachan, Nina Roussile, Kartini Shastry, Lise Vesterlund, and the participants in MIT's labor and behavioral lunches. This project was 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 resesarch funds, and the Administration for Children and Families (ACF) of the United States (U.S.) Department of Health and Human Services (HHS). The latter was as part of a financial assistance award (Grant #: 90PD0314) totaling $25,000 (28 percent) funded by ACF/HHS and $65,178 (72 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 pre-registration for the two experiments can be found here: https://www.socialscienceregistry.org/trials/9592 and here: https://www.socialscienceregistry.org/trials/11806. IRB approval for both studies was obtained under MIT's Committee on the Use of Humans as Experimental Subjects as protocols 2201000547 and 2307001048.

1 Introduction

Minority-group workers know discrimination is pervasive, have experiences that could be explained by discrimination, and often infer they have been discriminated against. In the last four decades, social scientists have documented widespread workplace discrimination (Goldin and Rouse, 2000; Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017; Neumark, 2018; Small and Pager, 2020). We know little, however, about the consequences of *perceived* discrimination because its effects are difficult to separate causally from those of discrimination itself.[1] Perceived discrimination is widespread, so it is important for understanding the impacts of discrimination in the labor market.[2] 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.

This paper describes two large, pre-registered experiments (N≈5000). The first experiment shows that perceived discrimination has quantitative importance for workers and firms; it also provides some evidence on how to reduce perceptions of discrimination. The second experiment tests the effects of common anti-bias hiring policies on *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. I focus on gender and racial discrimination in a white- and male-dominated domain (science), and oversample women and racial minority men.

The novel design element of both experiments is that workers are randomly assigned to be evaluated under procedures with varied potential to discriminate. When non-promoted workers learn how they were evaluated, I also truthfully vary what workers know about who was previously promoted under the same procedure. Then, I measure perceived discrimination and observe subsequent work behavior. I show that perceived discrimination is quantitatively important by focusing on the effects of treatment on worker performance, retention, and future labor supply. Specifically, I observe worker performance in required scientific proofreading tasks and the number of optional paragraphs they proofread—which I call retention—and elicit incentivized reservation wages for future job opportunities.

I also introduce a new measure of perceived discrimination that circumvents concerns with other methods. Prior surveys prime workers by asking directly about discrimination or differen-

[1] An exception is Gagnon et al. (2022), discussed below.

[2] 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 3,240 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 A17).

tial treatment (e.g. Kessler et al., 1999; Goldsmith et al., 2004; Mukerjee, 2014). 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 for them to be promoted. I define perceived discrimination as a worker's open-ended response mentioning demographic information, which correlates strongly with more explicit questions asked at the end of the survey.

In the first experiment (N≈2,400), 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; non-promoted workers are the experimental sample of interest. 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.[3]

I focus on the two manager arms to study the effects of perceived discrimination. Before starting the easier job, non-promoted workers see how they were evaluated and this generates 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.

Perceived discrimination affects retention and performance in ways that would exacerbate racial and gender gaps. Learning that one's manager knew 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. There are subsequently large effects on worker behavior. Being in the non-blind manager arm reduces the number of optional tasks completed by 3 percent (p=0.05). These effects are driven by white and racial minority women, for whom evaluation by a non-blind manager increases the probability of quitting as soon as they are able by 59 percent and lowers the chance of finishing the job by 8 percent (p=0.06, 0.04). There are also important, heterogeneous effects on performance in tasks required to receive the basic payment: being in the non-blind manager arm has no effect on white women's performance, worsens racial minority women's performance, and improves racial minority men's performance.

Perceived discrimination also reduces future labor supply, primarily due to direct disutility

[3]There was no observable discrimination. The algorithms selected workers at similar rates regardless of race and gender, and race and gender are not predictive of performance conditional on the algorithms' prediction. I could not control managers' decisions, but they did not appear to discriminate against women or racial minorities.

from future interactions rather than lower expected wages due to future discrimination. Workers in the non-blind manager arm have 9 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) psychological mechanisms, e.g. disutility from interacting with a biased manager. Only 25 percent of the 9 percent reduction is due to lower beliefs about the likelihood of promotion in the future. Workers are also 9 percent (p=0.09) less likely to share with their manager in a dictator game and have 18 percent (p=0.04) higher willingness to pay to be able to choose a different manager to work with in a collaborative task.

Since perceived discrimination worsens retention and some workers' performance, firms face financial incentives to reduce perceived discrimination and so should care about how their hiring procedures are perceived. Reducing actual bias may not reduce *perceived* discrimination, as reasoning about whether one has experienced discrimination is a complex problem. Consider the care with which social scientists design and implement the audit studies that document discrimination at scale (e.g. Kline et al., 2021). Individuals only receive few, noisy signals about why they do not receive certain opportunities and must draw their own conclusions about whether discrimination played a role (Jones et al., 2016; Doering et al., 2023). People struggle to interpret even simple probabilistic signals (Benjamin, 2019). Psychological and social complexities might additionally affect reasoning about discrimination since it is rooted in prejudice, oppression, and stereotypes.

In the second part of the paper, I test how firms might reduce perceived discrimination and its effects among their employees. First, I test whether anti-bias hiring procedures—using unbiased algorithms to rank candidates and blinding decision-makers to demographic information—affect perceived discrimination. In my setting, workers receive a strong signal that there may be racial or gender bias since the majority of previously-hired workers are white men. I also (truthfully) randomly vary whether workers see previously-hired workers who are all or two-thirds white men to understand how changes in minority-group representation affect worker perceptions of these policies. The experimental manipulations mimic ongoing structural changes in the labor market: the growing ubiquity of algorithms in managerial decision-making (Cowgill, 2020; Jarrahi et al., 2021) and pressure on companies to de-bias and diversify their workforces (Chang et al., 2019; Brecheisen, 2023; Fath, 2023; Gallup Inc, 2023).[4] Finally, I consider whether algorithmic evalua-

[4]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 are unavoidably opaque, undermining this potential. Companies' increasing emphasis on diversity, equity, and inclusion may successfully change perceptions, or fail or backfire if they seem inauthentic or ineffective. Flory et al. (2021) show that job ads emphasizing company 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 penalties and less likely to hire diverse candidates. Blinding decision-makers to demographic information can

tion can mitigate the *effects* of perceived discrimination. Algorithmic bias is generated by statistical processes, not explicit prejudice, which may change how perceived discrimination affects behavior (Bigman et al., 2023).

The second experiment (N≈2,700 new workers) tests how anti-bias hiring policies affect perceived discrimination. 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. Perceived discrimination is the primary outcome, and I do not measure effort or performance in a work task—the sample is workers who are not "hired," rather than not "promoted." Everything else about the labor market is the same as the first experiment.

I test whether minority-group representation—i.e. the *outcomes* of these procedures rather than the *inputs*—affects perceptions using additional random variation from the first experiment. Some workers are randomly assigned to managers who previously promoted three white men while others are assigned to managers who promoted two white men and someone else. Among those evaluated by the demographic-blind algorithm, workers are cross-randomized to learn previously-promoted workers' race and gender or not; there is similar variation in whether these are all or two-thirds white men.

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," but increasing minority-group representation effectively does so—an example of the representativeness heuristic possibly affecting belief formation (e.g. Kahneman and Tversky, 1972; Bordalo et al., 2021).[5] Specifically, when workers see that only white men were previously hired and know their manager knew their race and gender, 52 percent perceive discrimination.[6] Blinding managers to race and gender only reduces perceived discrimination by 11pp (20 percent, p-value=0.04) 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 9-11pp (20 percent, p<0.03). These results suggest that transitions to algorithmic decision-making or efforts to de-bias managers can, at best, only moderately reduce perceived discrimination if the outcomes of those procedures do not change. If these procedures increase representation of minority groups, however, they may be more effective

effectively improve representation of minority groups (e.g. Goldin and Rouse, 2000; Fath, 2023), but a recent survey of HR practitioners suggests that less than 20 percent of firms are employing this strategy (Fath and Zhu, 2021).

[5]This is conditional on the hiring policies used in the experiment—increasing representation via other policies, like affirmative action, may have other effects.

[6]This is higher than above because all previously-hired workers are white men and women (who are more likely to perceive discrimination) make up a higher fraction of the second experimental sample.

at reducing perceived discrimination: seeing one woman or racial minority man among the three previously-selected workers reduces perceived manager and algorithmic discrimination by 40-50 percent (p<0.01).

Since workers do perceive algorithmic discrimination, a final question is whether algorithmic decision-making could mitigate the effects of perceived discrimination on worker behavior. In the first experiment, workers evaluated by the algorithm are cross-randomized to see previously-promoted workers' profiles with or without demographics. This generates differences in perceived algorithmic discrimination in the sample in which I observe workers' subsequent performance in the work task. Seeing that a demographic-blind algorithm previously promoted mostly white men—which increases perceived discrimination by 17pp—reduces performance by 5-6 percent (p=0.07, 0.09) but does not affect retention. When workers anticipate future discrimination, perceived algorithmic discrimination also has similarly negative effects on future labor supply as perceived manager discrimination.

In sum, this paper shows that rectifying the disparate impacts of discrimination will require addressing worker perceptions due to both material and psychological costs. Furthermore, employers could improve equity and efficiency by reducing perceived discrimination, but eliminating such perceptions will likely be unattainable as long as minority groups remain under-represented. 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. 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; Glover et al., 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 objectively neutral. Prior work therefore understates the overall effects of a discriminatory labor market.

This paper's second contribution is to disentangle the effects of perceived discrimination into (i) changes in returns to effort or the wage value of work opportunities versus (ii) psychological mechanisms like retaliation or 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). I also build on a theoretical and experimental literature on

the effects of *anticipated* discrimination on human capital investment due to changes in incentives (Lundberg and Startz, 1983; Coate and Loury, 1993; Fryer et al., 2005; de Haan et al., 2017; Dianat et al., 2020) 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; Angeli et al., 2023; Avery et al., 2023).[7] These papers do not identify the mechanisms behind these effects. 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 in and out of jobs and increases willingness to take a pay cut to avoid future discrimination, due to both a direct utility cost and lower expected wages.

Most closely related in the literature on discrimination is contemporaneous work by Gagnon et al. (2022). In an online experiment, they show that explicitly attributing wage inequality 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. Rather than directly informing workers that discrimination occurred, I study a situation in which workers learn about hiring procedures and their results, and possibly infer discrimination. I provide evidence on what policies might reduce perceived discrimination and use a novel measure to document the prevalence of perceived discrimination. I also show that inferred perceived discrimination affects retention, performance, and future labor supply.[8]

Finally, an extensive literature in computer science, law, and economics shows that algorithms can either exacerbate or mitigate human bias (e.g. Raghavan et al., 2020; Kleinberg et al., 2020; Obermeyer et al., 2019; Kleinberg et al., 2018). Consistently, there is mixed evidence on how people perceive algorithmic versus human fairness and little work on perceived bias specifically.[9] 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 jobs when they will be evaluated by an algorithm rather than managers because they anticipate less discrimination (Avery

[7] Other work shows that Black workers reporting previously experiencing discrimination search for jobs more widely and that controlling for perceived discrimination explains racial differences in labor supply and job satisfaction (Goldsmith et al., 2004; Mukerjee, 2014; Pager and Pedulla, 2015).

[8] There are three 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 is stereotypically gender and racial discrimination (proofreading and summarizing scientific articles), versus their neutral context (copying randomly-generated strings of numbers and letters). Finally, I also document effects on cooperation with and retaliation against managers who workers perceive are biased.

[9] 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).

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 have negative effects on subsequent behavior.

The paper proceeds as follows: Section 2 describes the setting and research design. Section 3 provides evidence on the first research question—the effects of perceived discrimination. Sections 4, 5, and 6 provide evidence on the effect of anti-bias policies on perceived discrimination, the role of minority-group representation, and whether algorithms can mitigate the effects of perceived discrimination, respectively. Section 7 discusses robustness and Section 8 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. This experiment overcomes these challenges by creating random variation in perceived discrimination. Importantly, it does so while minimizing 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 (though in this setting, none observably did so). 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—who are primarily white men due to their 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 effects of actual discrimination (though there is none), (ii) all information workers have about the jobs, selection procedures, and other workers, and (iii) the effect of not being promoted. The main limitation is that, unlike most workplaces, workers and managers have short, impersonal "interactions." The implications for generalizability are ambiguous: with extended interactions, workers may perceive more discrimination if they take it more personally or less discrimination if they know more context. Effects on behavior may be larger in a more interactive setting if they arise due to retaliation or distrust, or smaller if there are higher stakes.

In both experiments, I over-sample workers likely to experience and perceive racial and gender discrimination. 2,080 workers make up the analysis sample in the first experiment, of which 50 percent are white women, 21 percent each are racial minority women and men, and 8 percent are white men. 2,527 workers make up the 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.[10] In both samples, the racial minority men and women are one-third each Asian, Black, and Hispanic. For comparison, the population of US workers *excluding white men* is 41 percent white women, 30 percent racial minority women, and 29 percent racial minority men (calculated from the 2021 American Community Survey, ACS).

The experimental sample includes people 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 by 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 on average more educated, but poorer and younger than a nationally representative sample. The fraction with a job outside Prolific is the same as the fraction of the US population in this age range that is in the labor force. Table 1 compares summary statistics from the experimental samples to their analogues from the nationally-representative 2021 ACS. Column 3 restricts the ACS sample to individuals ages 18-71, the minimum and 99th percentile of age in the experimental sample, and column 4 further restricts to employed women and racial minority men. The experimental sample is more likely to have a college degree, but also more likely to have household income betweem $20,000-$60,000 and less likely to earn above $160,000 than either

[10] 3,240 [3,960] workers were recruited to the baseline survey for the first [second] experiment. 120 [360] are dropped from 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 under any procedure regardless of their random assignment and make up the final analysis sample.

representative sample.

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. This experiment primarily tests whether perceived discrimination affects worker behavior, and provides some additional evidence on how to reduce perceived discrimination.

I refer to the second experiment 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; the primary goal of this experiment is to test whether anti-bias hiring policies affect *perceptions* of discrimination.

2.2.1 The job

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 in the *Science Journal for Kids* 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 (*Nature, Science,* etc.) 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.[11]

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 given 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, however, are told only they are paid for each paragraph in which they do "a good job" and that it was expected that about half of the paragraphs would qualify for payment.

[11] 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.

The scientific proofreading job was chosen for several reasons. First, women and racial minority men are under-represented and stereotypically under-perform in most scientific fields. Second, 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. Third, unlike other science- or math-based survey tasks, performance might be subjective, and as just described, workers knew only that they would be paid per paragraph that they did a "good job" proofreading. The effects of perceived discrimination could therefore have operated through workers anticipating further discrimination in payment decisions. Empirically, however, workers anticipated the use of objective performance metrics when asked at the end of the survey how they thought "a good job" was determined.

What did workers know about the jobs? At the beginning of the experimental survey, workers are told they are working on a team that proofreads scientific articles, and that 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 Appendix Figure A1). [12] They then learn they were assigned to the easier job.

2.2.2 Design overview and background

Appendix Figure A2 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 Appendix Figure A3). 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. Then, the experimental workers see who their manager or the algorithm previously promoted when they learn about their own evaluation (see Section 2.2.3).

The managers. The managers, all white men,[13] are recruited from Prolific and tasked with

[12] At the end of the survey, 68 percent of workers report they would have preferred 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.

[13] This serves two purposes: to best represent cases in which a women or racial minority man might be the most likely

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 (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' self-identified race, gender, and an avatar that workers created in the baseline survey (see below). Baseline performance is shown as 1-5 stars indicating the quintile of workers' average quiz scores and education is in three broad categories: no college degree, college degree, and 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 also evaluate workers with similar quiz scores as those they evaluated in the historical sample (see below). 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. Appendix B provides details about manager recruitment and randomization.

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). Ties are broken randomly. It was trained on a sample of 500 workers recruited from MTurk using Cloud Research's pre-approved participant pool. Like the experimental sample, women and racial minority men are over-represented in the training sample. Appendix C provides more detail.

2.2.3 Experimental design

This section describes the promotion experiment design, using the vocabulary from the previous section (i.e. *algorithm*, *demographic-blind manager*, *non-blind manager*, and *previously-promoted workers*, which refers to those promoted in the historical sample). Appendix Figure A4 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 and report how well they thought they did on each.[14] Workers know their performance on the quizzes could affect

to feel discriminated against and to minimize noise in the experiment caused by variation in manager characteristics (only 54 participated in the study).

[14] A spelling quiz of 10 scientific words, 15 grammar questions, and 11 science questions. The difference between workers' performance and their (incentivized) beliefs about performance is a measure of confidence, which could predict perceived discrimination if motivated reasoning causes some workers to attribute rejection to discrimination rather than their performance (Heidhues et al., 2019). This does not seem to be the case (Appendix Figure A18).

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,"[15] 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 with average baseline quiz scores within one quintile of each other. Each group is evaluated by two randomly-assigned managers, one demographic-blind and one non-blind, and the algorithm. Both managers had previously been randomly assigned to evaluate workers in the same two quiz-score quintiles in the historical sample, so workers with low scores see previously-promoted workers with similar scores to themselves and can infer they had a chance at getting the harder job. Otherwise, low-scoring workers would mainly see top-scoring workers being promoted and would be unlikely to attribute decisions to discrimination.[16]

Within each group, 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 are balanced across treatment arms on baseline characteristics (Appendix Table A1).

Workers that are not promoted under their randomly-assigned promotion procedure are offered the easier job and the 80 percent who take it up become the experimental sample (N=2,397). They are all offered the same Prolific task (i.e. same wage and job description) and only learn about their evaluation *after* opting in. Attrition after starting the survey was uncommon, and similar across all treatment arms (Section 7). Promoted workers are offered the harder job.

Information about non-promotion and evaluation procedures. The timeline of the experimental survey is shown in Appendix Figure A5. Workers are invited to take the experimental survey at least three weeks after taking the baseline survey. Upon starting the survey, workers learn about the two jobs and that they were not promoted (i.e., they were assigned to the easier, lower-paying job). 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.

[15] Workers made their avatars following the procedure shown in Appendix Figure A19. They do not know what the avatars will be used for or that they will be seen by other participants. They therefore should *not* have chosen their avatars strategically, indeed, workers largely choose avatars that align with their self-reported race and gender in expected ways (Appendix Figure A20).

[16] Workers were told "[your manager/the algorithm] evaluated workers with similar quiz scores as you." As intended, workers at all performance levels were similarly likely to perceive discrimination, conditional on being in the higher or lower quintile in their group, though workers in the fifth quintile in the 4-5 quintile group were more likely to perceive discrimination than those in the higher quintile in the 3-4 and 2-3 quintile groups (Appendix Figure A21, Panel A). Unsurprisingly, workers in the higher of the two quintiles of their group are more likely to perceive discrimination. Workers with more education were also more likely to perceive discrimination (Panel B).

In the manager sample, workers learn that managers was recruited from Prolific, reviewed example paragraphs, evaluated worker profiles, and are paid based on the performance of the workers they promote and so should be trying to promote the best workers. They also see their manager's profile (avatar, age, education, gender, race, and tenure on Prolific). Thus, 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 (and thus perceived discrimination) comes next: they are told 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. Appendix Figure A6 shows the implementation.

In the non-blind manager arm, workers see profiles that include avatars, average quiz-score quintiles as 1-5 stars, education (no college degree, college degree, or more than college degree), self-identified gender (Man, Woman, or Non-Binary/Other), and self-identified race/ethnicity.[17]

In the demographic-blind manager arm, the profiles show only quiz-score stars and education. This controls for any effect of feeling misjudged for not having a college degree, misunderstanding what the stars represented, or not being promoted.

Thus, "treatment" in the manager sample is learning that one's manager knew race and gender 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, all 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 in the same format as the non-blind manager arm: they include avatars, stars, education, gender, and race. The other half see profiles in the same format as the demographic-blind manager arm: just stars and education. Appendix Figure A7 shows the implementation. 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. This is random due to the random pairing of workers with managers or groups of historical-sample workers jointly evaluated by the algorithm.

[17] Only 6 of 2397 participants identified as non-binary; they are grouped with women in results that split by gender as the "not male" category.

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 in their evaluation.[18]

To my knowledge, no other study has attempted to identify individuals who felt that they had been discriminated against in a recent interaction.[19] This measure aims to capture that feeling without priming workers to think about discrimination or their own race and gender. Other, more explicit questions that identify perceived discrimination—asked at the end of the survey and described below—correlate highly with the main measure (Appendix Figure A8).

Observed work outcomes. Workers are required to proofread six paragraphs (spending a 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 standard psychological scale that elicits current emotional states.

Incentivized survey outcomes. I elicit reservation wages via a multiple price list: workers indicate for various wage schedules whether they would like to participate in a future round of the survey at the given wage schedule 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: (1) if they would be evaluated by the same procedure as in the experiment and (2) if their assignment would depend only on whether they had the highest baseline quiz scores. After each multiple price list, they report what they think

[18] Age was not communicated directly but was somewhat observable through avatar hair color and baldness. This was pre-registered as the main measure of explicit perceived discrimination. The free response measure was coded as mentioning demographics independently by two MIT undergraduates and a PhD psychologist who professionally codes qualitative data and was hired on Upwork. The externally-generated variables used in the analysis are highly correlated with variables generated by the author. The full coding scheme for this and other text-based variables is in Appendix H.

[19] 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, usually "ever" or in the past year.

is the probability that they will be promoted if they participate in the future round with the given evaluation procedure.

In the manager arms only, I similarly elicit incentivized reservation wages to participate in a collaborative task with their manager from the experiment and their willingness to pay to be able to choose a different manager to work with in the cooperative task.[20] Workers in the manager arms also play a dictator game. They know they may be randomly chosen to receive a $20 thank-you bonus for their participation and choose how much they would want to share with their manager if they are selected, in which case their choice will be implemented. These outcomes measure workers' willingness to share and work cooperatively with their manager, proxying for e.g. willingness to stay late at work or take on extra tasks for their manager.

Self-reported outcomes. Workers next 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 similar methods as the main measure (Appendix H). 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. These more direct measures of perceived discrimination are highly correlated with the main measure (Appendix Figure A8). 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.[21]

2.3 The hiring experiment

The hiring experiment tests how firms' hiring policies affect whether workers perceive discrimination. Specifically, it identifies the effects of using an algorithm versus a manager and the effect of

[20] Workers summarize the paragraphs used in the harder version of the proofreading job, managers provide edits, and workers make revisions. They earn a base wage that the manager cannot influence, but worker-manager pairs who provide some of the best summaries will earn a bonus, and the manager has discretion over how it is split.

[21] All of these "explicit" measures of perceived discrimination require the worker to think about or describe an experience of discrimination. I also observe implicit measures, which are differences in the number of stars that workers report thinking they would have needed to be promoted compared to other demographic groups. The signs of the treatment effects on these variables are consistent with the explicit measures, but generally imprecise (available upon request).

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 (Section 2.1); the only sampling difference is that white women are more heavily over-represented, due to constraints on the number of racial minority participants available who had not already participated in the promotion experiment. The not-hired workers take an experimental survey without any proofreading job, and the key outcome is perceived discrimination.

3,960 *new* workers are recruited to 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 (Appendix Figure A9). Workers are randomly assigned to be evaluated by a manager or an algorithm and cross-randomized to arms where the decision-maker can use 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 A2).

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. The rest 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 interacts each baseline performance and education variable in the demographic-blind algorithm with the eight race $\times$ gender groups and predicts performance. Most algorithms would not explicitly use race and gender information, but a realistic hiring algorithm would use sufficient predictors to be able to approximate race and gender. I aim to proxy this type of black-box algorithm in a setting without large amounts of data. The non-blind algorithm does not observably discriminate—it meets the same calibration criteria as the demographic-blind algorithm (Appendix C)—and in fact, would hire more racial minorities than the demographic-blind algorithm (Appendix Table A3).

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 evaluation procedures 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 a reason outside the standard codes (Appendix H).

In the hiring experiment, workers also answered comprehension questions about the procedures at the end of the survey. Approximately 80 percent of workers in all manager arms correctly identified the managers' payment structure and approximately 70 percent of workers in all algorithm arms correctly identified that the algorithm predicts performance. 88-98 percent of workers correctly identified 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 un-incentivized survey section. 75 and 80 percent of workers answered the question correctly in the promotion and hiring experiments, respectively.[22] There is no differential attention across treatment arms and the main results are robust to dropping the workers who failed the attention check (Section 7).

3 The effects of perceived discrimination

3.1 Estimation

To understand the effects of perceived discrimination, 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.[23] Appendix D provides details on all

[22] Workers were asked to select one column for both rows in a matrix with two Likert-scale questions.

[23] In the main specification, X includes baseline quiz scores, education, income, age, family 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 had one fewer star) and ten exhaustive and mutually-exclusive indicators for how the workers' education compared to the education of the previously-promoted workers

outcome and control variables. Standard errors are robust to heteroskedasticity.

One threat to the causal interpretation of this specification as the effects of interventions that change only *perceptions* of discrimination is that different evaluation procedures could have promoted different types of workers. Since their real decisions were implemented, this could generate differences between the treatment arms among workers who are not promoted. To eliminate this threat, 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—i.e. the workers not promoted or hired under any procedure. In the promotion experiment, this is 2,080 workers of 2,317. In the hiring experiment, this is 2,527 workers of 2,680. In practice, the different procedures do not appear to select systematically different workers (Appendix Table A3) and the workers who are selected under at least one procedure appear similar to those who are never selected (Appendix Table A4).

The choice of controls and analysis sample do not affect the main results. They are also robust to multiple hypothesis testing and using randomization inference (Section 7).

An instrumental variables (IV) specification. I also present IV specifications in Appendix E, instrumenting for the main measure of perceived discrimination with random assignment. These effectively divide the ITT estimates of the effects on behavior by the "first stage" effect on perceived discrimination rates. 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, there may be a certain degree of perceived discrimination that has effects, which 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, and thus implies 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 and represent a lower bound on the average effects of perceived discrimination caused by a particular treatment.

3.2 Manipulation check: Treatment affects perceptions

Workers' perception of discrimination strongly depends on whether they learn that their manager knew demographic information and previously promoted mostly white men. Figure 1 plots rates of perceived discrimination in each manager arm estimated using Equation 1 for the full sample and separately for women, racial minority men, and white men.[24]

they saw. The controls in H were not pre-registered; the results are the same excluding these controls (Section 7).

[24] Heterogeneity by race and gender was pre-registered, along with heterogeneity by past experiences of discrimination, beliefs about the prevalence of discrimination, and confidence. I focus on gender and race as they most systematically

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 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).[25] Being in the non-blind manager arm increases perceived discrimination by 37pp (se=3pp) for women and 20pp (se=4pp) for racial minority men. Appendix Figure A10 plots rates in each treatment arm for each of eight race $\times$ gender cells; women generally perceive discrimination at similar rates regardless of race as do racial minority men.[26]

Appendix Table A5 shows parallel results using the secondary measures of perceived discrimination. There is always a large, significant effect, and patterns across demographic groups are the same for all measures, though the prevalence in the demographic-blind manager group and magnitude of treatment effects vary.

3.3 Results: Effects of perceived discrimination

3.3.1 Retention and performance

Learning that one's manager knew race and gender and previously promoted mostly white men lowers retention—persistence in the work task—and performance, particularly for women, though it also improves some workers' performance.

Retention. Workers must proofread six paragraphs to be paid their participation wage, and can then choose to proofread up to eighteen. Learning that one's manager knew race and gender and previously promoted mostly white men—i.e. being in the non-blind manager arm—causes workers to proofread 0.5 fewer paragraphs (3 percent, se=0.25, p=0.05) and reduces the probability that they proofread all eighteen paragraphs by 4.9pp (6 percent, se=2.5pp, p=0.06) (Table 2).

predict perceived discrimination, but multivariate regressions that include all pre-registered characteristics are in Appendix Figure A18.

[25] 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 (27 and 13 percent, se=2.4pp and 2.6pp), respectively (Appendix Table A10). Note that 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 by 18pp (se=2pp, Appendix Table A5).

[26] Why are there such stark gender differences? Gender discrimination could have been more salient than racial discrimination in this experiment, as the avatars' shirt colors differed by gender but race was only partially communicated in the avatar via skin tone. However, among racial minority women, the majority of those who perceive discrimination report that they think they would have been hired if they had a different race *or* if they had a different gender; of those who only mention one or the other, about twice as many perceive only racial discrimination. Aksoy et al. (2023) similarly find that women are more likely to anticipate discrimination on the basis of some other negatively-discriminated characteristic than men (in their case, sexual orientation).

Intensive effort and performance. I identify causal effects of treatment on effort and performance in the first six required paragraphs, before treatment causes 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). It also does not affect performance—the number of correct or incorrect highlights, or earned bonus (Table 3). This masks substantial heterogeneity, however, which is discussed below.

Total earnings. If the bonus was a piece-rate per paragraph, the negative retention effect would correspond to the same reduction in total earnings. In this experiment, however, the bonus 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, perceived discrimination decreases total earnings for some but increases total earnings for others. The overall effect of being in the non-blind manager arm on total earnings is zero (Appendix Figure A11). Combined with the negative effect on retention, this implies that those who stay experience a *positive* treatment effect of perceived discrimination on earnings in the later paragraphs, canceling out the negative earnings effect on those induced to quit early. While this effect cannot be quantitatively separated from selection effects in who chooses to stay, heterogeneous treatment effects by race and gender provide a sense of magnitudes.

Heterogeneity by race and gender. The negative effects on retention are driven by women, and racial minority women experience a negative performance effect in the required paragraphs, where selection cannot play a role. In contrast, racial minority men experience positive performance effects. Figure 2 plots the effect of being in the non-blind manager arm on the cumulative percent quit, bonus earned, and correct and incorrect highlights separately by race and gender.

White and racial minority women drive the negative effects on retention. White women proofread 0.65 (4 percent, se=0.32) fewer paragraphs and are 3.9pp (70 percent, se=2.1pp, p=0.07) more likely to quit as soon as they are able. Racial minority women proofread one (6 percent, se=0.56, p=0.07) fewer paragraph and are 12pp (16 percent, se=5.7pp) less likely to complete all eighteen paragraphs. In contrast, racial minority men proofread 0.81 *more* paragraphs, though the estimate is not statistically significant. The effects on retention for men and women are significantly different from each other at a less than 5 percent level (Appendix Table A6, Panel A. Panel B pools all women, with similar results).

Racial minority men and women's performance is also affected. Racial minority men *correctly* highlight 1.3 (10 percent, se=0.47) more mistakes in the required paragraphs, whereas racial minority women *incorrectly* highlight 0.8 (30 percent, se=0.41) more non-mistakes. They also highlight about 5 percent more words correctly, but the effect is not statistically significant (Appendix Table

A7, Panel A). Thus, women and men respond differently when they perceive discrimination—on average, white and racial minority women disengage and do worse, while men try to prove themselves, in this case, successfully. Racial minority women may similarly try to prove themselves by highlighting more words, but do so unsuccessfully, or disengage and simply highlight more words less thoughtfully. The positive performance effect for racial minority men in the required paragraphs is significantly higher (p=0.0*3*) from the zero effect for women when white and racial minority women are pooled together (Panel B).[27]

Gender differences in confidence and competitiveness (Niederle and Vesterlund, 2007; Coffman, 2014; Bordalo et al., 2019) could underly this result, though other factors may also be at play. Consistently, though not statistically significant, there is a weakly positive effect of being in the non-blind manager arm on racial minority men's psychological well-being—they are less upset, discouraged, annoyed, and anxious—and a weakly negative effect for racial minority women—they are more annoyed and less motivated, though also less anxious (Appendix Figure A12).[28]

In summary, perceived discrimination overall negatively affects retention, with important heterogeneous effects on performance. Some workers may react by trying to prove themselves to biased managers, but only some do so successfully. Even if future evaluations are race- and gender-neutral, perceived discrimination may 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.[29]

3.3.2 Future labor supply

Perceived discrimination also 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.[30]

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. 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

[27] 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 2).

[28] There are no effects on self-efficacy or job satisfaction (Appendix Tables A11 and A12).

[29] I focus on heterogeneity by race and gender as it yields the most consistent patterns of the pre-registered heterogeneity analysis. Appendix F looks at other pre-registered dimensions of heterogeneity. The negative effect of perceived discrimination on retention is most negative for workers with below-median confidence, after accounting for racial and gender heterogeneity. The stark gender heterogeneity remains after accounting for confidence.

[30] 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 the outcomes discussed in this section were also measured in the hiring experiment; this was pre-registered as a replication of the promotion experiment results. Indeed the results are very similar; they are in Appendix I.

always twice the lower. In other words, each row of the multiple price list reflected two wages, e.g. "Earn $0.20 per paragraph if not promoted and $0.40 per paragraph if promoted." 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.[31] I use a 50 percent probability because the average worker in the control group believes this is the likelihood of future promotion.[32]

Thus, the effect on this formulation of reservation wages could be driven either by beliefs about future discrimination or psychological mechanisms like anticipated disutility. I decompose these channels in Section 3.3.3.

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 cent) if workers would be evaluated by the same manager in the future (Table 4). Those in the non-blind manager arm are also willing to pay an additional 1 cent per paragraph (se=0.5 cents) to be evaluated by a cutoff rule in baseline performance in which there is no potential for future discrimination or interaction with your manager, measured as the difference in reservation wages if they would be evaluated by their manager or the cutoff rule. This willingness to pay is economically meaningful: it is 4 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 (Table 5). Being in the non-blind manager arm increases workers' willingness to pay to be able to choose a different manager in a cooperative job by 4.1 cents per paragraph (18 percent, se=2 cents). Workers in the non-blind manager arm are 5pp (10 percent, se=3pp, p=0.09) more likely to share $0 with their manager of their possible $20 bonus, and overall share about 26 cents (9 percent, se=23 cents, p-value=0.28) less with their manager (Table 5).[33]

Applied to a work context where employees and managers interact on a regular basis, this

[31] Sample sizes are slightly smaller for these variables as I drop observations in which workers make inconsistent choices indicating they do not understand the MPL elicitation method, which is standard in this type of analysis. Appendix G provides more information and shows that the effects on retention and performance are similar in this restricted sample.

[32] 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). Appendix Figure A22 plots the CDFs of beliefs by treatment group; a large share of workers guess 50 percent but there is wide variation.

[33] This "dictator game" measures distributional preferences and retaliatory preferences if one has already interacted with the recipient. Workers whose managers did not see demographic information shared on average $2.80 (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 affects 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.

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.[34]

3.3.3 Mechanisms

The effects on willingness to interact with and generosity towards managers 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.[35]

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. Next, I describe a framework to disentangle these effects, map the framework into the experimental design, and then describe results.

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 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

[34] In terms of race and gender, being in the non-blind manager arm increases all workers' willingness to pay to avoid their manager and causes retaliation (Appendix Table A13). That said, managers may react more negatively to women or racial minorities who are less cooperative or generous. Appendix F looks at other pre-registered dimensions of heterogeneity. The positive effect of perceived discrimination on willingness to pay to choose one's own manager is largest for workers who report experiencing discrimination at work in the past in the baseline survey.

[35] Theory and lab experiments focus on returns to effort (Lundberg and Startz, 1983; Coate and Loury, 1993; Fryer et al., 2005; de Haan et al., 2017; Dianat et al., 2020); field experiments show effects in high-stakes settings but do not separate (seemingly 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., 2023).

not affect beliefs about the likelihood of 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, there are two other ways that perceived discrimination could affect expected utility and thus reservation wages. First, beliefs about future promotion may depend on D independently of $\hat{\sigma}$: $\hat{\pi} = \pi(\hat{\sigma}, D)$. This could be because perceived discrimination changes e.g. 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 $\hat{\pi}$. In words, workers may dislike discrimination or dislike interacting with a biased manager or an employer who hires biased managers, even if they do not anticipate future discrimination. The anticipation cost is increasing in both $\hat{\sigma}$ 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 and changes in e.g. trust or self-confidence each explain about half of the effect of perceived discrimination on beliefs about the likelihood of promotion. Second, the resulting change in expected utility from wages explains only 25 percent of the overall effect on reservation wages. The effect on anticipated disutility $C(\hat{\sigma}, D)$ explains the rest. It seems to derive about equally from (i) disutility from anticipated discrimination or interactions with biased managers and (ii) disutility from continuing to work for the employer.

First, I estimate the effects on the likelihood of future promotion, separately when future discrimination is possible (when one will be evaluated by the same manager, i.e. $\hat{\sigma} \geq 0$) and when it is not (when one will be promoted 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 no future discrimination is possible. Being in the non-blind manager arm lowers beliefs about the likelihood of future promotion by 2.7pp (6 percent, se=1.2pp). About 40 percent of this effect, however, can be explained by changes in beliefs about the likelihood of promotion when no future discrimination is possible: 1.5pp (se=0.7pp) of the effect of perceived discrimination on beliefs about the likelihood of future promotion 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 multiple price list assuming that workers had the *average*

worker's belief about the likelihood of promotion. Now, the high and low wages in the multiple price list are multiplied by workers' *own* belief about the probability that they would be assigned to each type of job.[36] This adjusts for any effect of perceived discrimination on beliefs about the probability of promotion, but 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. When future discrimination is possible, effects on reservation wages constructed in this way can come from $\frac{\partial C}{\partial \hat{\sigma}}$ or from $\frac{\partial C}{\partial D}$; when future discrimination is not, the effects isolate $\frac{\partial C}{\partial D}$.

Being in the non-blind manager arm increases reservation wages for future evaluation by the same manager by 1.7 cents per paragraph (a 6.5 percent increase, se=1 cent, p=0.10) using this construction of reservation wages. This is 74 percent of the total effect on reservation wages (2.3 cents per paragraph) when beliefs about promotion can also play a role (Figure 3, top panel). When workers will be evaluated by the cutoff rule, treatment increases reservation wages by 0.9 cents per paragraph after accounting for beliefs (though this is statistically insignificant). Taking the point estimates at face value, this means that about half of the overall effect on $C(D, \sigma)$ is therefore due to disutility from anticipated discrimination or continued interactions, and the other half 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 7.9pp (12 percent, se=2.7pp), are interested in more jobs with tasks assigned like this one by 5.3pp (9 percent, se=2.9pp, p=0.07), and are interested in doing the harder job in the future by 6.3pp (12 percent, se=2.9pp) (Appendix Table A8).

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

In fact, these costs are also the most likely explanation for the negative effects on retention discussed in the previous section. While proofreading, workers were not aware of the specifics of the future work opportunities or that they would possibly interact with their manager again, and they did not think that their manager would be involved in determining their bonus payment.[37]

[36] E.g. if a worker thinks there is a 20 percent chance of being promoted, and they switch from not wanting and wanting to be evaluated between the piece rates 30 and 35 cents per paragraph for the lower-paying job (and correspondingly 60 cents and 70 cents per paragraph for the higher-paying job), then their reservation wage would be calculated as $0.8 \times 32.5 + 0.2 \times 65 = 39$ cents per paragraph, whereas a worker with a higher expected probability of promotion but the same switching point would have a higher estimated reservation wage.

[37] 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

Thus, it is unlikely that perceived discrimination affected worker behavior through the channel of changes in expected wages, so the large negative effects of perceived discrimination on retention for women are most likely due to similar retaliation or avoidance mechanisms.

A psychological channel that could explain the documented effects of perceived discrimination on performance is by activating *stereotype threat* (e.g. Spencer et al., 2016; Liu et al., 2021) for women or the opposite *stereotype challenge* (e.g. Alter et al., 2010) for men. This is not something the experiment was designed to isolate from other psychological channels but could be considered an underlying channel through which perceived discrimination affected the psychological costs of effort.

Together, the results from the promotion experiment show that perceived discrimination has quantitatively important effects on worker behavior and subsequent outcomes. These effects would exacerbate gender gaps conditional on race and racial gaps among women. 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. They could improve retention, some workers' performance, and worker-manager cooperation by doing so. The next sections test how firms might achieve this goal.

4 The effects of anti-bias policies on perceptions

4.1 Estimation

The hiring experiment tests whether procedural changes that reduce the likelihood of actual discrimination—blinding managers to demographics and using unbiased algorithms—mitigate perceptions of discrimination. I estimate the following:

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

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

arm mentioned "manager" or "person/people reviewing" 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, and many workers also mentioned a cutoff rule in the number of correct answers.

gender. All other variables are the same as in the previous analysis (Section 3.1). Now, I focus on one outcome: Y_i is an indicator for whether workers perceive discrimination.

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, χ_1 (χ_2) when the decision-maker is demographic-blind (non-blind) and workers do 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 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. In the main analysis, the above regression is estimated only on those workers (a random subset, as in the promotion experiment). [38]

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 (2,527 workers of 2,680). Again, in practice, this does not seem to be a concern (Appendix Tables A3 and A4). Standard errors are robust to heteroskedasticity. The results do not depend on any of these specification decisions (Section 7).

4.2 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 only white men were previously promoted*. Figure 4 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

[38] Results for the full sample are similar but conflate perceptions of procedures with minority-group representation. (Appendix Figure A23). The procedures are differentially likely to hire entirely white men in the historical sample. This exacerbates differences in perceived discrimination between the algorithm and manager arms compared to the results in the main text (by depressing rates of perceived discrimination in the manager arms) and reduces differences between the demographic-blind with avatars arm and non-blind manager arm (by depressing rates of perceived discrimination by more in the non-blind manager arm than in the demographic-blind manager arm).

percent of workers evaluated by a manager and 63 percent evaluated by an algorithm perceive discrimination, an 11pp difference (se=5pp). This difference is similar when decision-makers do not know demographics (a 9pp difference, se=4pp, Figure 4).

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 differences are similar using the secondary measures of perceived discrimination (Appendix Figure A13).

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—in other words, they see previous decisions that "look like" discrimination, and convincing them otherwise is hard. In fact, while 50-60 percent of workers in the other four arms correctly identify the information that their manager or the algorithm knew when making hiring decisions, only 7 and 16 percent of workers do so when they are evaluated by a demographic-blind manager and algorithm, respectively, if they see previous hires' avatars, race, and gender (Appendix Figure A14). This is *not* due to (differential) inattention or general confusion (Section 2.4).

Instead, they are more likely to be incorrect precisely because they are highly likely to think the decision-maker knew their race and gender (Appendix Figure A14). 95 percent of workers correctly identify that the non-blind manager and algorithm knew race and gender, but 87 and 70 percent of workers *incorrectly* think so when they are evaluated by a demographic-blind manager and algorithm, respectively but see that the previously-promoted workers are all white men. That said, I cannot separate disbelief or motivated reasoning from differential forgetfulness or confusion about this particular decision input.[39]

At least some workers perceive algorithmic discrimination even if they know the algorithm only uses past performance and education as inputs. In the demographic-blind arms that did not see avatars, race, and gender for previous hires, 21 and 7 percent of workers thought that the manager and algorithm used demographic information, respectively. The difference is likely due to workers' seeing *managers'* demographics in all manager treatment arms. Thus, there is no significant difference in the effect of seeing the previous hires' avatars, race, and gender on whether workers think demographic-blind decision-makers used demographic information between the manager arms and the algorithm arms (66pp versus 64pp, Appendix Figure A14). Recall that workers are 9pp (21 percent) *more* likely to perceive discrimination by an algorithm than a manager when both are demographic-blind. Together, the results suggests that workers are more likely to know that

[39]The information about decision-makers' inputs and who was previously promoted were both communicated visually, so should have been similarly salient (Appendix Figure A24).

the algorithm didn't use demographic information but perceive discrimination all the same. One explanation is that they understand that algorithms can discriminate without using demographics as inputs but managers cannot. In general, however, workers do not understand the implications of what they are told about the algorithm's design, whereas they do understand the manager's incentives (Appendix Figure A15). So instead, general distrust or misunderstanding of the algorithm may make them more likely to jump to conclusions when they see what "looks like discrimination."

5 The effects of minority-group representation

5.1 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 that two white men and someone else was previously promoted. There is truthful, random variation in the demographic composition of the previously-promoted workers that each worker in the experimental sample saw (Appendix Figure A4).

In the manager arms, workers were randomly assigned to managers who had made different choices in the historical sample. In the non-blind manager arm of the promotion experiment, 46 percent of workers were randomly assigned to managers who had previously promoted three white men and 54 percent were assigned to managers who had promoted two white men and someone else. Workers who see that two versus three white men were previously promoted have similar observable characteristics (Appendix Table A9).[40] In the demographic-blind arm, this was 40 percent and 60 percent, respectively, but note that these two groups of workers did not actually see anything different about the previously-promoted workers.

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. Recall that workers in the algorithm arm were cross-randomized to either learn the demographics of previously-promoted workers or not. When workers learned previously-promoted workers' demographic information, 33 percent saw three white men and 67 percent saw two white men and someone else. Workers who see that two versus three white men were previously promoted have similar observable characteristics (Appendix Table

[40] One concern is that the fraction of white men among previously-promoted workers could co-vary with other manager characteristics, but this is not reflected in the data and there was minimal heterogeneity in manager profiles (Appendix Table A14).

A9). I estimate the following:

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

where $3WM_i$ is an indicator for being evaluated by a manager who previously promoted three white men, 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. All other variables are the same as previously(see Section 3.1), and Y_i indicates perceived discrimination.

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. Note that this analysis was not pre-registered (see Section 7).

5.2 Results

More representation of women and racial minority men among previously-promoted workers substantially reduces perceived discrimination. Seeing one minority-group member among the three previously-promoted workers reduces rates of perceived discrimination by 17pp (40 percent, se=3pp) in the non-blind manager arm and by 15pp (50 percent, se=5pp) in the algorithm arm when workers see previously-promoted workers' demographics (Figure 5).

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 oneself reduces perceived discrimination from 44 to 34 percent relative to seeing three white men (a 10pp difference, se=4pp), and seeing one minority-group member from one's *own group* reduces perceived discrimination by an *additional* 18pp (se=5pp). The pattern is the same in the algorithm arm (Appendix Figure A16).

Thus, worker beliefs about discrimination are much more responsive to signals from decision

outputs—who was selected—than they are to signals about inputs—information about how the decision was made. When a decision "looks like discrimination," workers pay little attention to decision inputs—a potential application of the representativeness heuristic affecting belief formation (e.g. Kahneman and Tversky, 1972; Bordalo et al., 2021). When previously-promoted workers include members of minority groups, especially one's own group, workers are much less likely to perceive discrimination even when there could have been discrimination.

Taken together, these results have several implications for settings where some demographic groups are severely under-represented. First, anti-bias policies—like transitioning to algorithmic screening or censoring demographic identifiers from resumes—may slightly reduce perceptions of discrimination, 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 also suggest that settings with more equal representation may have less perceived discrimination. Future 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.

6 The effects of perceived algorithmic discrimination

Since implementing algorithmic hiring procedures cannot fully eliminate perceived discrimination (Section 4), including if it improves minority-group representation among previously-promoted workers (Section 5), a final related question is whether algorithmic hiring procedures can mitigate the *effects* of perceived discrimination. People are less morally outraged by depictions of algorithmic bias than human bias and are less likely to attribute algorithmic bias to prejudice when it does occur (Bigman et al., 2023); if the effects documented in Section 3 are due to perceived *prejudice* rather than perceived *discrimination,* they may not arise when workers perceive algorithmic discrimination.

6.1 Estimation

To understand the effects of perceived *algorithmic* discrimination, I focus on intent-to-treat (ITT) specifications that are analogous to those in Section 3.1. Restricting to workers randomly assigned to be evaluated by the algorithm in the promotion experiment, I now compare outcomes between workers who were cross-randomized to see previously-promoted workers' race, gender, and avatar versus those who did not. Now, T_i in Equation 1indicates seeing the demographics of previously-promoted workers; β is the effect of learning that the demographic-blind algorithm previously

promoted mostly white men.

6.2 Manipulation check: Treatment affects perceptions

Again, worker perceptions 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.[41]

Learning that the algorithm previously promoted mostly white men increases perceived discrimination rates from 0 to 17 percent (se=2.6pp), even though workers know the algorithm did not access demographic information (Figure 6). The effect is 22pp for women (se=3.4pp) and 8.5pp (se=5pp) for racial minority men.[42]

6.3 Results: Effects of perceived algorithmic discrimination

The effects of perceived algorithmic discrimination seem to be negative, but differ in some ways from the effects of perceived manager discrimination.

Retention. Learning that the demographic-blind algorithm previously promoted mostly white men does not statistically significantly affect retention. The point estimates across all three outcomes (number of paragraphs proofread, proofreading more than the required six paragraphs, and finishing all eighteen paragraphs) are positive, but the confidence intervals are wide and the ITT effect of the algorithm sub-treatment cannot be statistically distinguished from the negative ITT effect of being in the non-blind manager arm on retention (Figure 7, Row 1).

Performance and earnings. Learning that the demographic-blind algorithm previously promoted mostly white men negatively affects performance in the first six required paragraphs (the number of words correctly highlighted and bonuses earned), and these effects are statistically significantly different from the ITT effect of being in the non-blind manager arm relative to the demographic-blind manager arm (Figure 7, Row 3). Learning that the demographic-blind algorithm previously promoted mostly white men lowers performance by about 5 percent.

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 unbi-

[41] Given high-profile articles about algorithmic discrimination (e.g. 'Amazon scraps secret AI recruiting tool that showed bias against women', Reuters), this was not obvious *ex ante*. It was possible that individuals would infer bias just from the word "algorithm."

[42] Results are similar using the secondary measures of perceived discrimination (Appendix Table A15). In turn, it reduces the share who cite the primary reason for their non-promotion being their education by 7pp (21 percent, se=3.4pp, Appendix Table A16) while it increases the share who cite discrimination as the *primary* reason by 9pp (se=2pp, Appendix Table A15).

ased cutoff, but these are not statistically distinguishable from the positive effects in the manager arms (Figure 7, Row 4). These smaller, less distinguishable effects may be due to the fact that there is no effect of learning that the demographic-blind algorithm previously promoted white men on beliefs about future promotion or anticipated discrimination (Figure 7, Row 5). This did not, however, replicate in the hiring experiment, where these beliefs and reservation wages were the only outcomes collected that related to worker behavior. Appendix I describes these results, which were pre-registered as a replication exercise. In the hiring experiment, perceived algorithmic discrimination did significantly increase anticipated discrimination in the future and reservation wages for future work (as did perceived manager discrimination). Thus, the anticipated psychological costs of algorithmic discrimination and manager discrimination may be similar.

In summary, the use of algorithmic hiring procedures does change the effects of perceived discrimination, though the effects are still generally 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), though many other things are different about perceived human versus algorithmic discrimination that may drive these differences. Other potential explanations are differences in *who* perceives discrimination and heterogeneous treatment effects, the fact that algorithms cannot themselves "benefit" from a worker's performance, or the fact that an algorithm cannot be "proven wrong."

7 Robustness

The results are broadly quite 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 G provides the details and results.

Multiple hypothesis testing and randomization inference. The statistical significance of my main results does not change when I (1) account for multiple hypothesis testing using a Romano-Wolf (2005; 2016) correction or (2) 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).

Attrition and attention. There is no differential attrition in either experiment that can explain my results for the effects of perceived manager discrimination on worker behavior, the effects of anti-bias hiring procedures or minority-group representation on perceived discrimination rates, or the effects of perceived algorithmic discrimination. There is one difference in the rate at which

workers failed the attention check (between workers who saw three white men versus two white men previously promoted in the demographic-blind algorithm arm of the promoted experiment); the specification charts described next show that the corresponding effect on perceived discrimination is robust to dropping workers who failed the attention check. Section 2.4 discusses overall experimental fidelity and shows that worker attention to and comprehension of the experiment was high overall.

Specification choices. The main results largely do not depend on the choice of control variables or sample definition. The effects of perceived manager discrimination and the effects of anti-bias hiring procedures and minority-group representation on perceived discrimination rates are unchanged if I further restrict the sample to workers who passed the attention check near the end of the survey 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 are less negative and are statistical 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 the lack of observable differences between the treatment arms. The results are robust to replacing the specifications with the exact pre-registered specification along two dimensions, using lasso-selected control variables (Chernozhukov et al., 2018), or dropping all control variables.

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 randomly-determined) in an open-ended text response. I use coded measures of whether they think the study is about discrimination (Appendix H) 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 fraction 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. I use the more conservative bound for the latter since this outcome was not incentivized.

The effects of perceived discrimination on worker behavior are unchanged by this exercise.[43]

[43]This is consistent with Mummolo and Peterson's (2019) result that workers do not change their behavior when they are told the purpose of a study.

The effects of blinding hiring procedures to demographics and the effect of minority-group representation on perceived discrimination all fall by about half when replacing the perceived discrimination measure with zero for workers who believe the study is about discrimination, but remain statistically significant.

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 (Appendix D shows exact implementations). 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 G discusses how these workers compare to those who did answer sensibly 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 from the pre-analysis plans; again, see Appendix G for details and smaller deviations.

The primary deviation from the pre-analysis plan 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 experiment 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.

8 Conclusion

This paper shows that addressing the disparate impacts of discrimination will require addressing worker perceptions in addition to discrimination itself. Perceived discrimination negatively affects retention and future labor supply, which would 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 going to be difficult to do so purely by implementing anti-bias procedures in contexts where minority-group workers are severely under-represented. Reducing biased decision-making itself is of course the right place to start. 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—can only mildly reduce perceived discrimination if the outcomes of those procedures—who is hired or promoted—remain the same and members of under-represented groups continue to see decisions that "look like discrimination." Increased diversity may be a more promising tool, and perceptions of discrimination after affirmative action or other diversity initiatives warrant additional future study. Implicit bias training seems not to be very effective at reducing bias (Chang et al., 2019)—is it so popular because it helps reduce perceptions of discrimination?

These results offer other interesting avenues for future research on perceptions of discrimination as well. Can workers be convinced that algorithms are unbiased? More generally, how can they be encouraged to trust information about de-biased decision-making? Finally, perhaps there will always be times when some workers feel they have been discriminated against. Can firms mitigate the negative effects of perceived discrimination by encouraging workers to report such concerns and taking them seriously? What psychological or behavioral interventions might help workers overcome the effects of perceived discrimination on their engagement at work 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.

Alter, Adam L., Joshua Aronson, John M. Darley, Cordaro Rodriguez, and Diane N. Ruble, "Rising to the Threat: Reducing Stereotype Threat by Reframing the Threat as a Challenge," *Journal of Experimental Social Psychology*, January 2010, *46* (1), 166–171.

Angeli, Deivis, Ieda Matavelli, and Fernando Secco, "Expected Discrimination and Job Search," *Working Paper*, 2023.

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.

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.

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 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 Demand," *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.

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.

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

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

__ **and Susan Zhu**, "Preferences for, and Familiarity With, Blinding Among HR Practitioners," *Working Paper*, 2021.

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. Pages: xi, 251.

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*, 2022.

Gallup Inc, "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.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack, "Overconfidence and prejudice," *Working Paper*, 2019. arXiv: 1909.08497.

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, *2019* (1), 18172.

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.

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.

Liu, Songqi, Pei Liu, Mo Wang, and Baoshan Zhang, "Effectiveness of Stereotype Threat Interventions: A Meta-analytic Review," *The Journal of Applied Psychology*, 2021, *106* (6), 921–949.

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.

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.

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 Soci-*

ology, 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.

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

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.

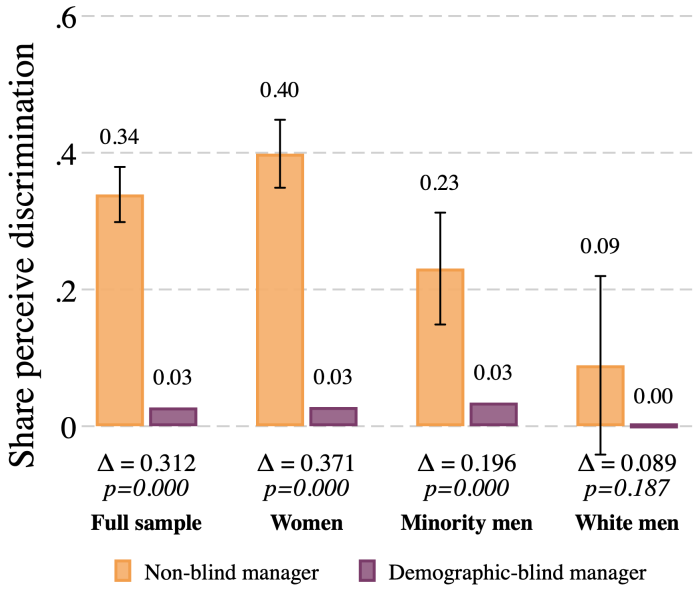
Small, Mario L. and Devah Pager, "Sociological Perspectives on Racial Discrimination," *Journal of Economic Perspectives*, 2020, *34* (2), 49–67.

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

Spencer, Steven J., Christine Logel, and Paul G. Davies, "Stereotype threat," *Annual Review of Psychology*, 2016, *67* (1), 415–437.

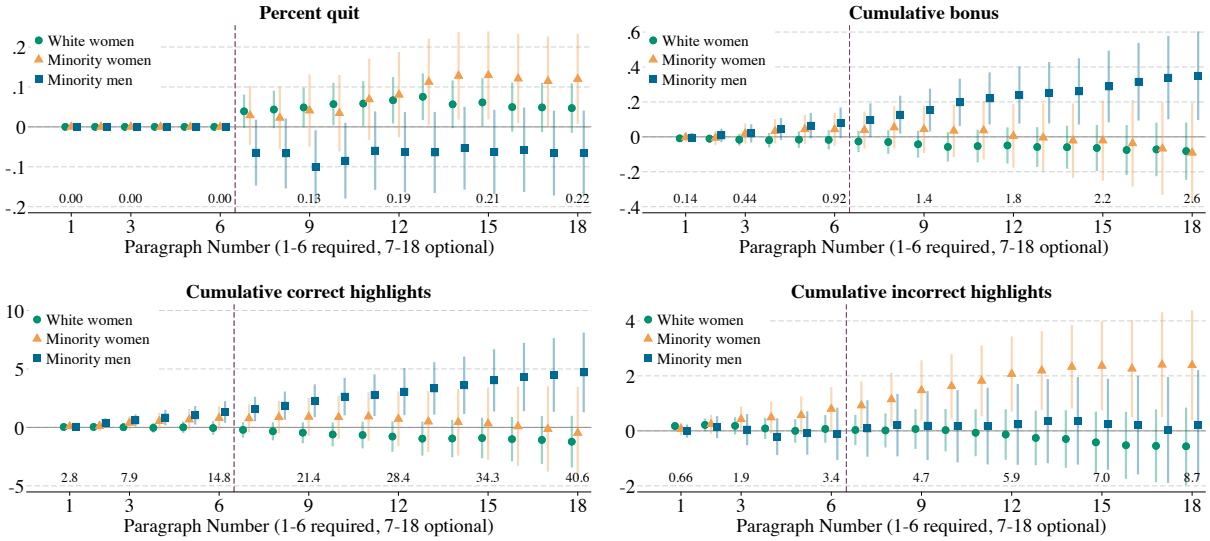
Zhang, Lixuan and Christopher Yencha, "Examining Perceptions Towards Hiring Algorithms," *Technology in Society*, 2022, *68*, 101848.

Figure 1: Perceived discrimination in the manager arms of the promotion experiment



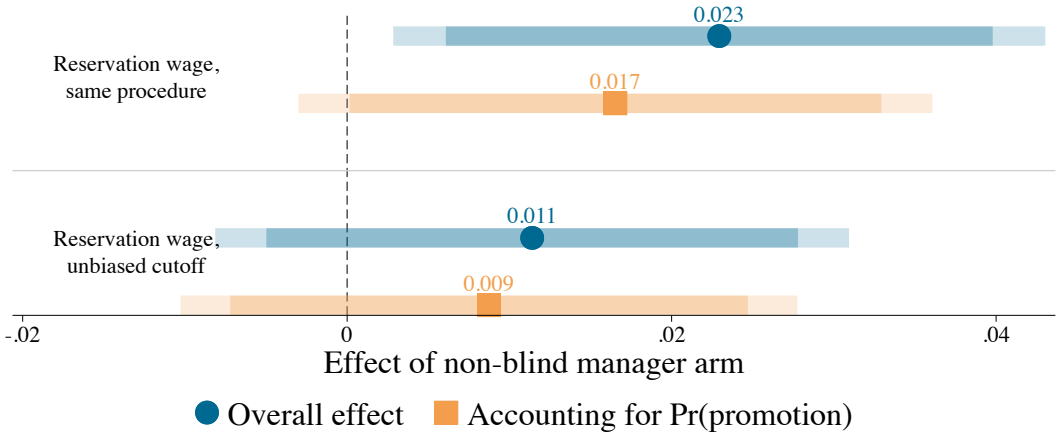
Note: This figure plots the share of workers in the experimental sample of the promotion experiment who perceive discrimination using the main measure (whether a worker mentions demographics in their response to the open-ended question, *"what needed to be different about your profile in order to be assigned to the harder task?"*) first for all workers and then separately for women (of all races), racial minority men, and white men in the two manager treatment arms. The plotted means are estimated from the coefficients estimated by Equation 1 separately for each subgroup. The figure restricts to workers who would not have been promoted under any promotion procedure. For breakdowns by race × gender, see Appendix Figure A10. 95 percent confidence intervals calculated with standard errors robust to heteroskedasticity are indicated by the black bars. The corresponding effects on the secondary measures of perceived discrimination are in Appendix Table A5.
[Click to return to Section 3 (Effects of Perceived Discrimination)]

Figure 2: Treatment effects on retention and performance by gender × race



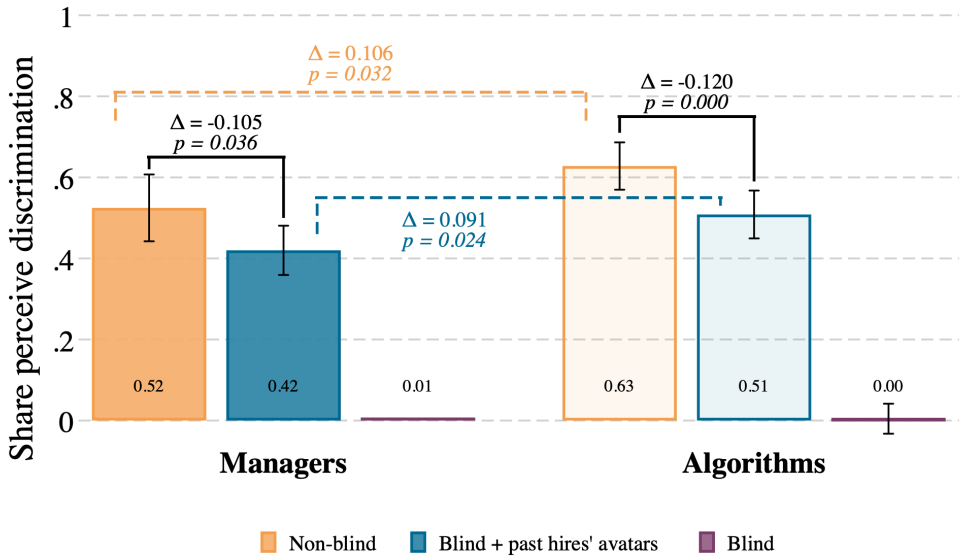
Note: This figure plots the effect of being in the non-blind manager arm of the promotion experiment on retention and performance measures, each cumulatively by the given paragraph. Equation 1 is estimated interacting the treatment indicator with indicators for each race/gender group; white men are dropped. 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3.1 (Effects on Retention and Performance)]

Figure 3: Decomposing the effects on reservation wages in the promotion experiment



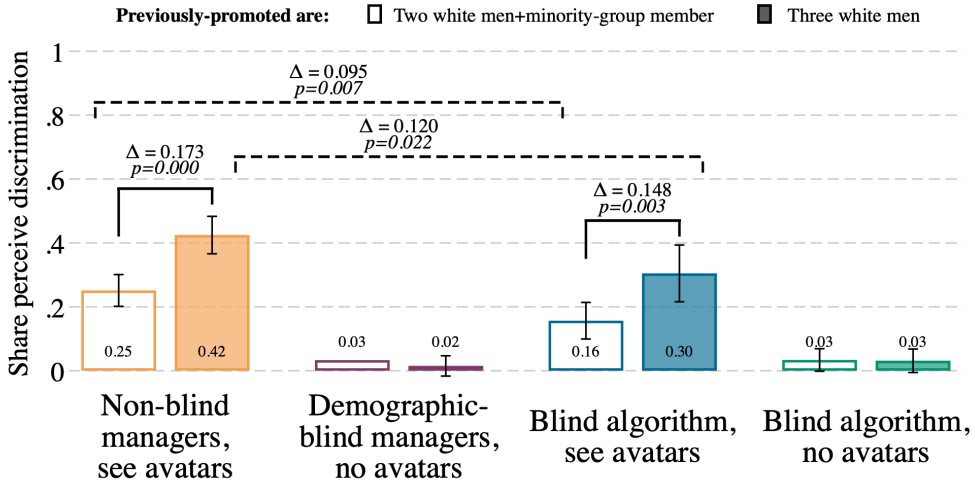
Note: This figure plots the effect of being in the non-blind manager arm relative to the demographic-blind manager arm of the promotion experiment on workers' reservation wages when they will be evaluated by the same manager as in the experiment (top panel) or by an unbiased cutoff rule in baseline quiz 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 2. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3.2 (Effects on Future Labor Supply)]

Figure 4: Perceived discrimination in the hiring experiment



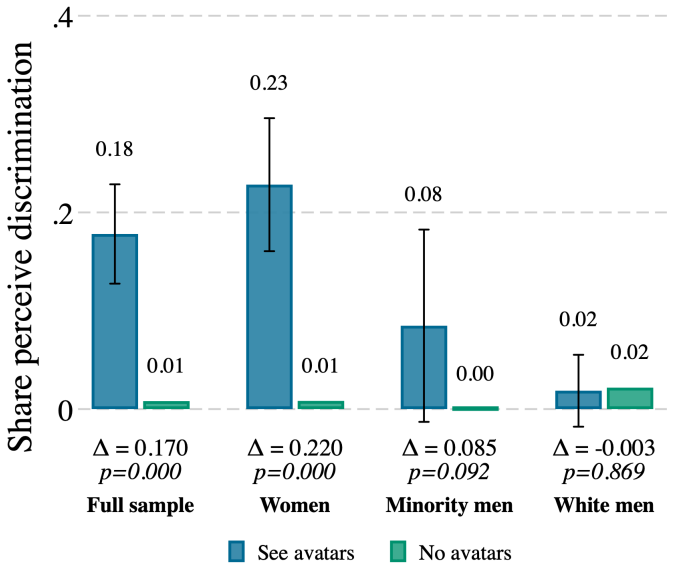
Note: This figure plots the share of workers perceiving discrimination in each treatment arm of the hiring experiment, using the main measure of perceived discrimination. The sample is restricted to workers who would not have been hired under any hiring procedure and workers who see three previous hires who are all white men (a random subsample within each treatment arm). Shares are calculated via regressions with the same controls as Table 2. p-values and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 4 (Effects of Hiring Policies)]

Figure 5: Effects of seeing one previously-promoted minority-group worker



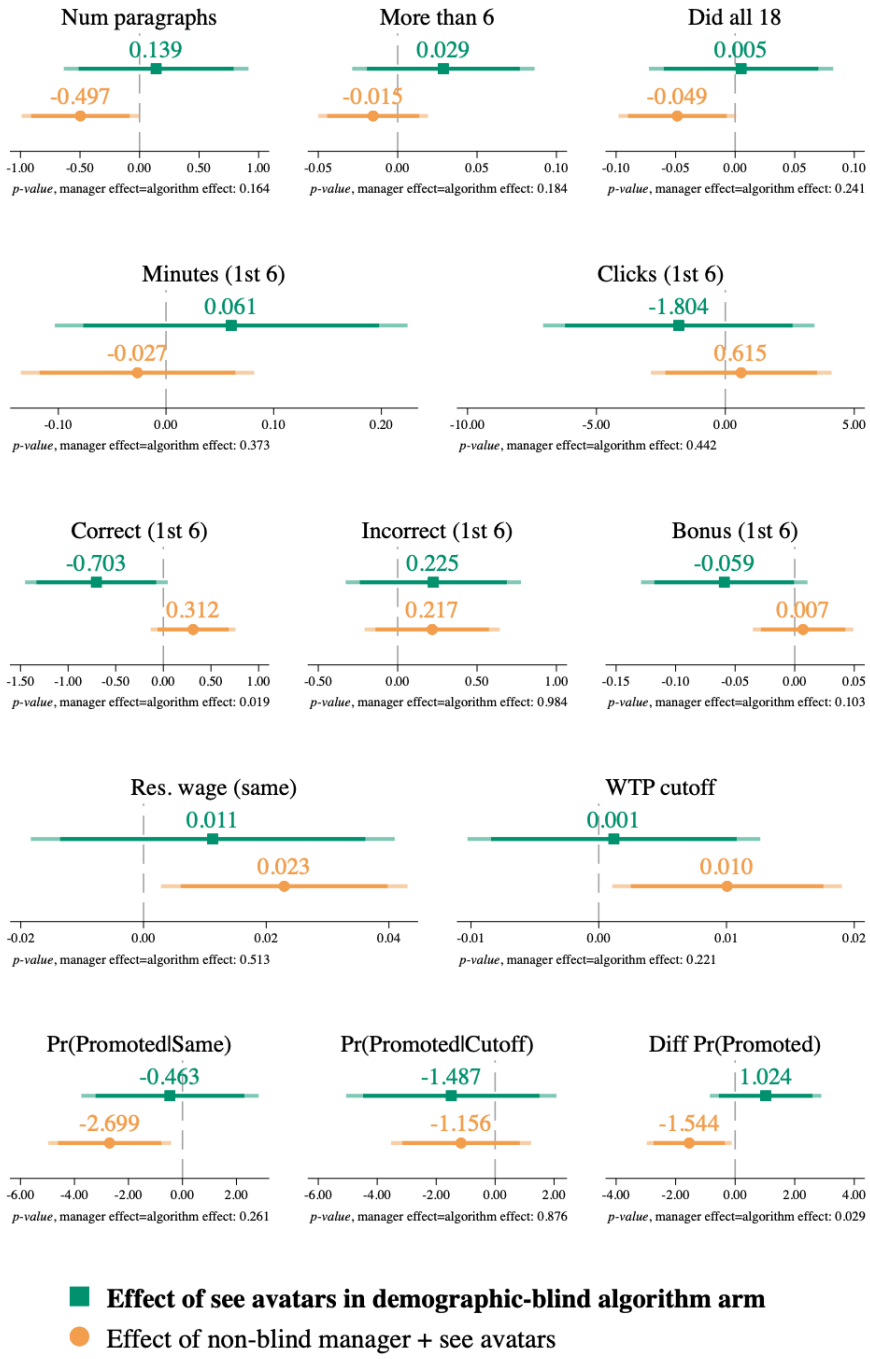
Note: This figure plots the share in each treatment arm of the promotion experiment perceiving discrimination by the main measure, separately by whether workers (would) see three white men previously promoted (40 percent of each arm) or not. Those who do not see three white men see two white men and 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 sample and controls are the same as Table 2. *p*-values and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 5 (Effects of Minority-Group Representation)]

Figure 6: Perceived discrimination in the algorithm arm of the promotion experiment



Note: This figure is analogous to Figure 1 but compares perceived discrimination rates in the two sub-arms of the algorithm treatment arm. The corresponding effects on the secondary measures of perceived discrimination are in Appendix Table A15.
[Click to return to Section 6 (Effects of Perceived Algorithmic Discrimination)]

Figure 7: 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 6 (Effects of Perceived Algorithmic Discrimination)]

Table 1: Summary statistics and comparison to ACS sample

	Experimental Samples		ACS 2021, ages 18-71	
	Promotion	Hiring	All	Subsample
	(1)	(2)	(3)	(4)
Male	0.29	0.13	0.50	0.31
Race:				
Asian	0.13	0.06	0.06	0.09
Black	0.14	0.12	0.12	0.17
White, Hispanic	0.14	0.10	0.18	0.27
White, Non-Hispanic	0.59	0.71	0.59	0.40
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	0.01	0.10	0.09
High school graduate	0.13	0.10	0.27	0.24
Some college but no degree	0.27	0.25	0.21	0.21
2 year college degree	0.09	0.10	0.09	0.09
4 year college degree	0.37	0.39	0.21	0.23
Professional or Masters degree	0.11	0.14	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	0.18	0.15	0.15
$40,000-$60,000	0.18	0.19	0.14	0.14
$60,000-$80,000	0.15	0.17	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	0.17	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	2527	2216940	1005688

Note: The samples in columns 1 and 2 are the analysis samples 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	More than 6	Did all 18
	(1)	(2)	(3)
Non-blind manager, see avatars	-0.497**	-0.015	-0.049*
	(0.252)	(0.018)	(0.025)
N	1387	1387	1387
Control mean	15.916	0.915	0.776

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. 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 levels indicated by *, **, and ***, respectively.
[Click to return to Section 3.3.1 (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, see avatars	-0.027	0.615	0.312	0.217	0.007
	(0.055)	(1.787)	(0.227)	(0.217)	(0.022)
N	1389	1389	1389	1389	1389
Control mean	4.957	36.804	14.817	3.422	0.919

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 levels indicated by *, **, and ***, respectively.
[Click to return to Section 3.3.1 (Effects on Retention and Performance)]

Table 4: Effects on future labor supply

	Res. wage (same procedure)	WTP unbiased procedure
	(1)	(2)
Non-blind manager, see avatars	0.023**	0.010**
	(0.010)	(0.005)
N	1338	1325
Control mean	0.265	-0.014

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. 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 3.3.2 (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	0.008	0.041**	0.049*	-0.255
	(0.015)	(0.020)	(0.029)	(0.233)
N	1349	1030	1383	1383
Control mean	0.341	0.223	0.522	2.857

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. 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 3.3.2 (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	-2.699**	-1.156	-1.544**
	(1.163)	(1.212)	(0.728)
N	1385	1385	1385
Control mean	47.219	48.491	-1.272

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 stated 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 levels indicated by *, **, and ***, respectively.
[Click to return to Section 3.3.2 (Effects on Future Labor Supply)]

Online Appendix: Perceived discrimination at work

Hannah Ruebeck

March 2024

Appendix Table of Contents

A Supplementary figures and tables

B Manager recruitment and manager task

C Training the algorithm

D Variable definitions

E Instrumental variables (IV) specification results

F Other heterogeneity

G Robustness analyses

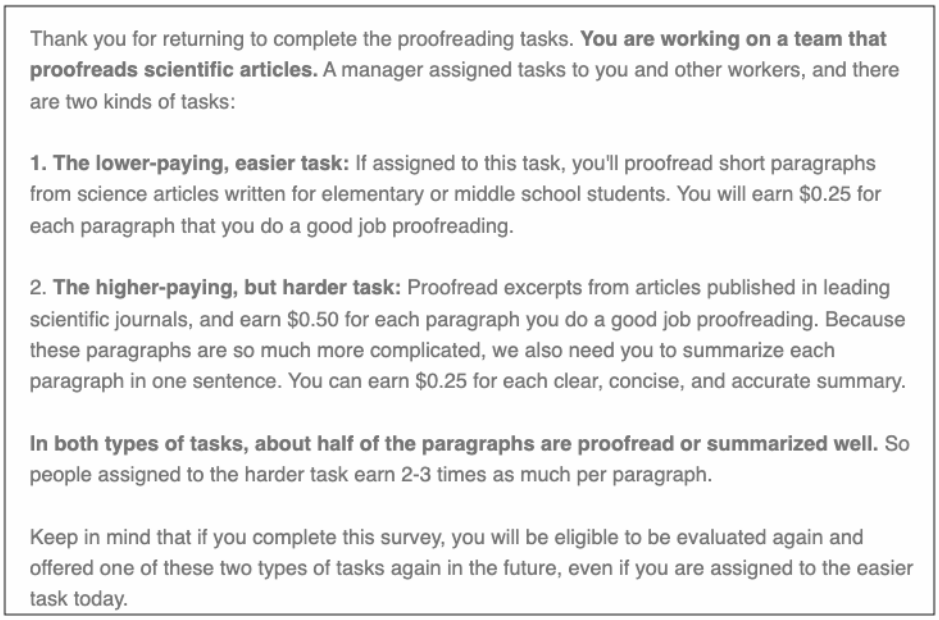
H Coding the text data

I Effects on future labor supply in the hiring experiment

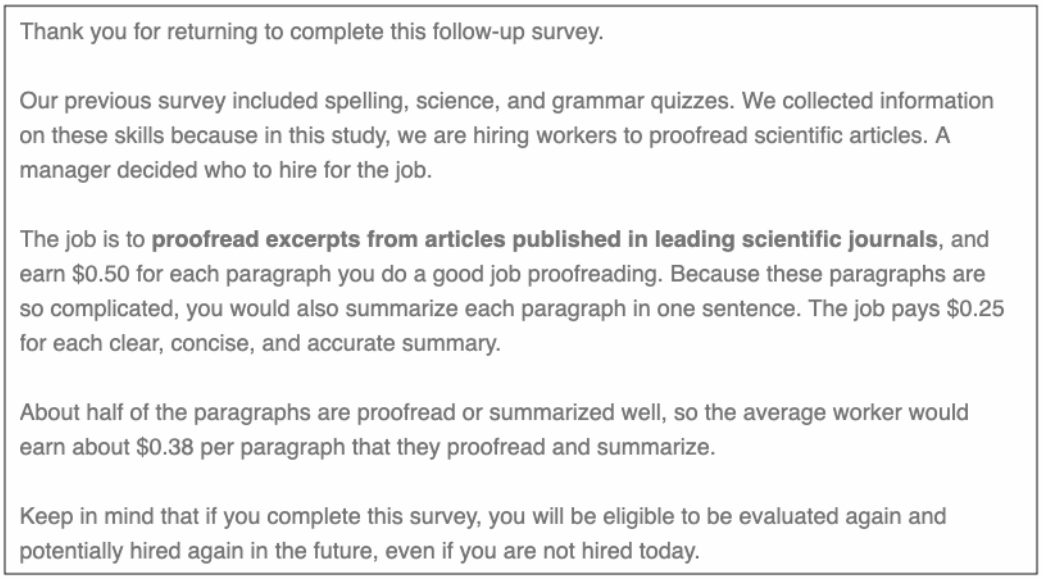
A Supplementary figures and tables

Figure A1: Information workers received about the job(s)

Panel A: Promotion experiment

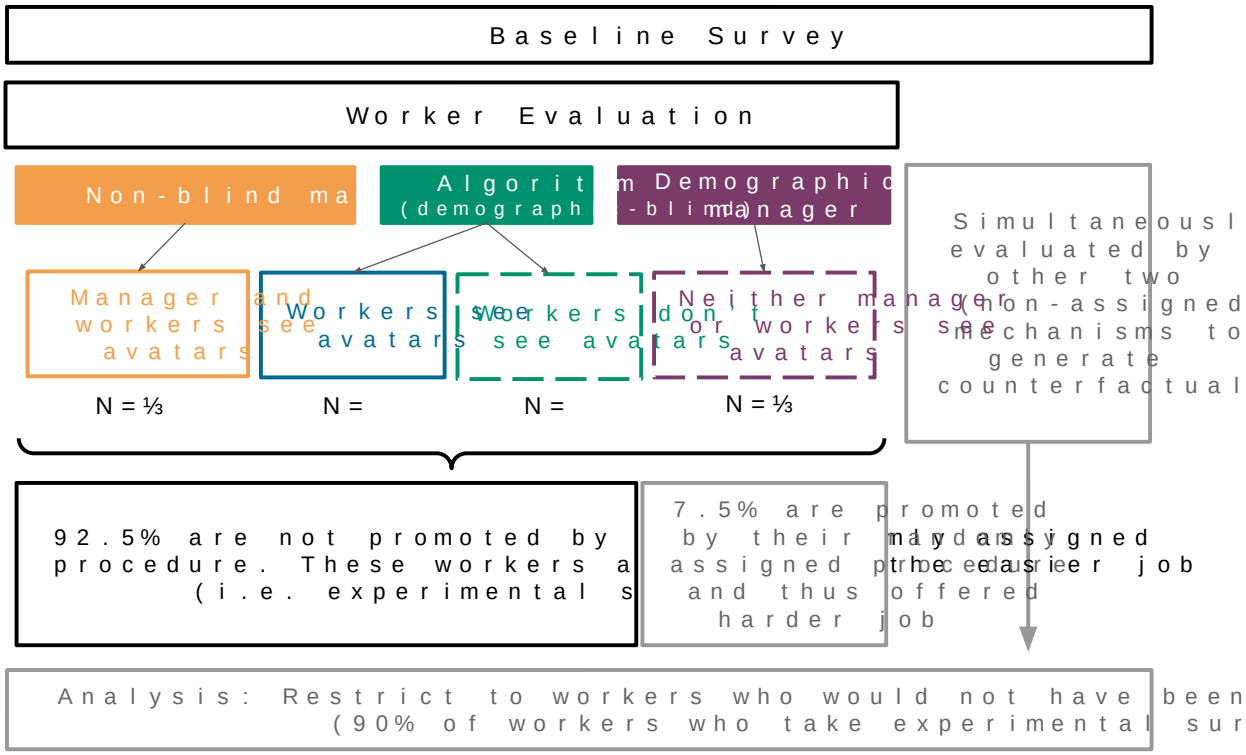


Panel B: Hiring experiment



Note: This figure shows the information that workers received about the easier, lower-paying ("non-promotion") job and the harder, higher-paying ("promotion") job in the promotion experiment (Panel A) or the job that they were not hired to do in the hiring experiment (Panel B).
[Click to return to Section 2 (Research Design)]

Figure A2: 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 and did not complete the survey afterwards to reduce costs. The design of the hiring experiment was the same except that workers were assigned to be evaluated by one of four procedures (see Figure A9). In addition, 97.5% of workers were not hired under their randomly assigned procedure.
[Click to return to Section 2 (Research Design)]

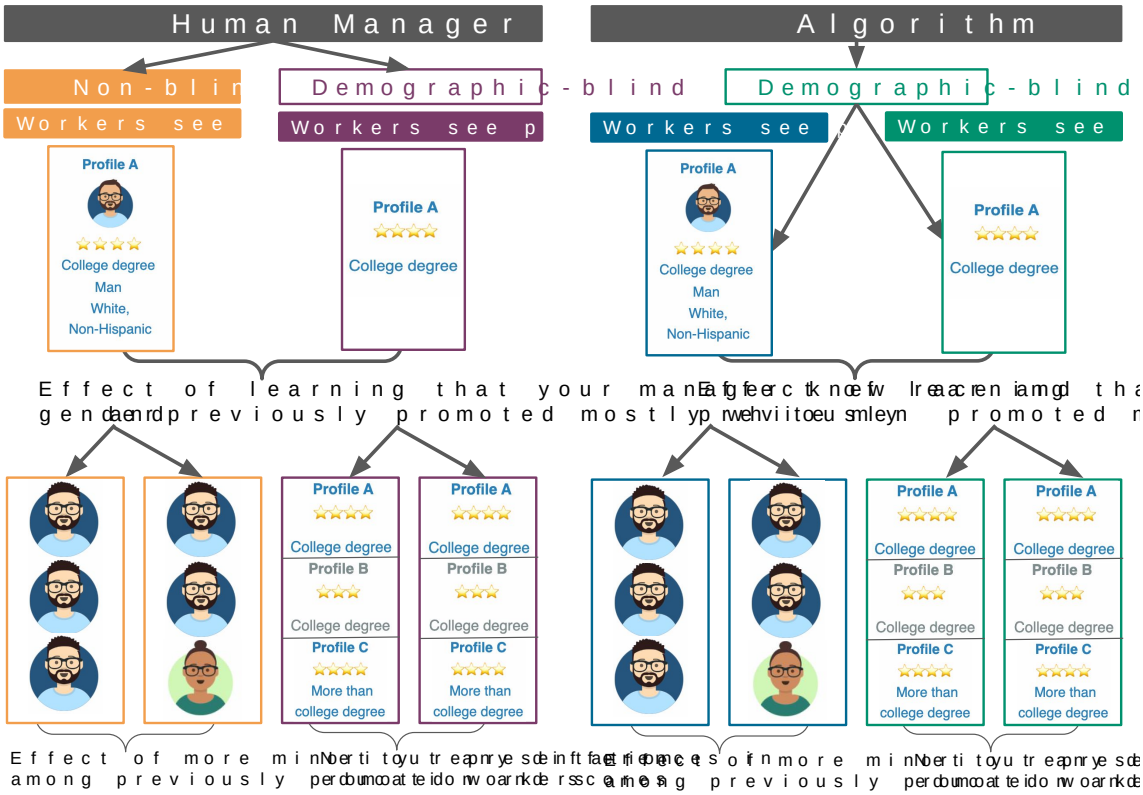
Figure A3: Timeline of study



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.

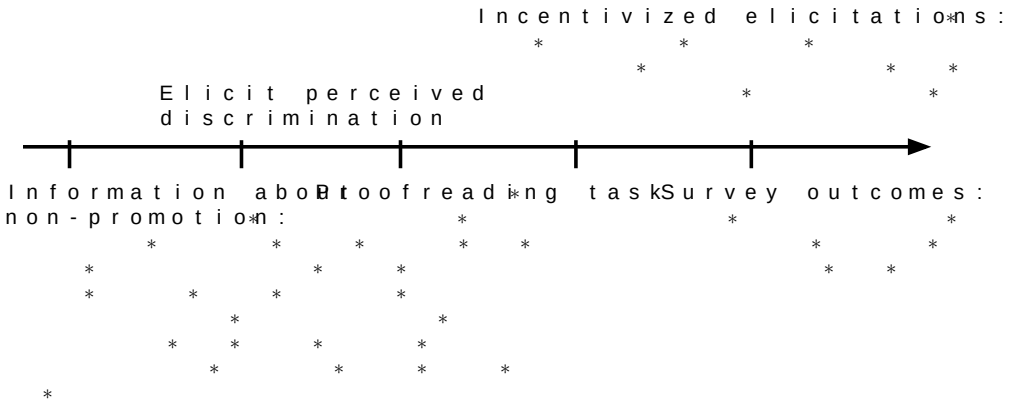
[Click to return to Section 2 (Research Design)]

Figure A4: Design of promotion experiment



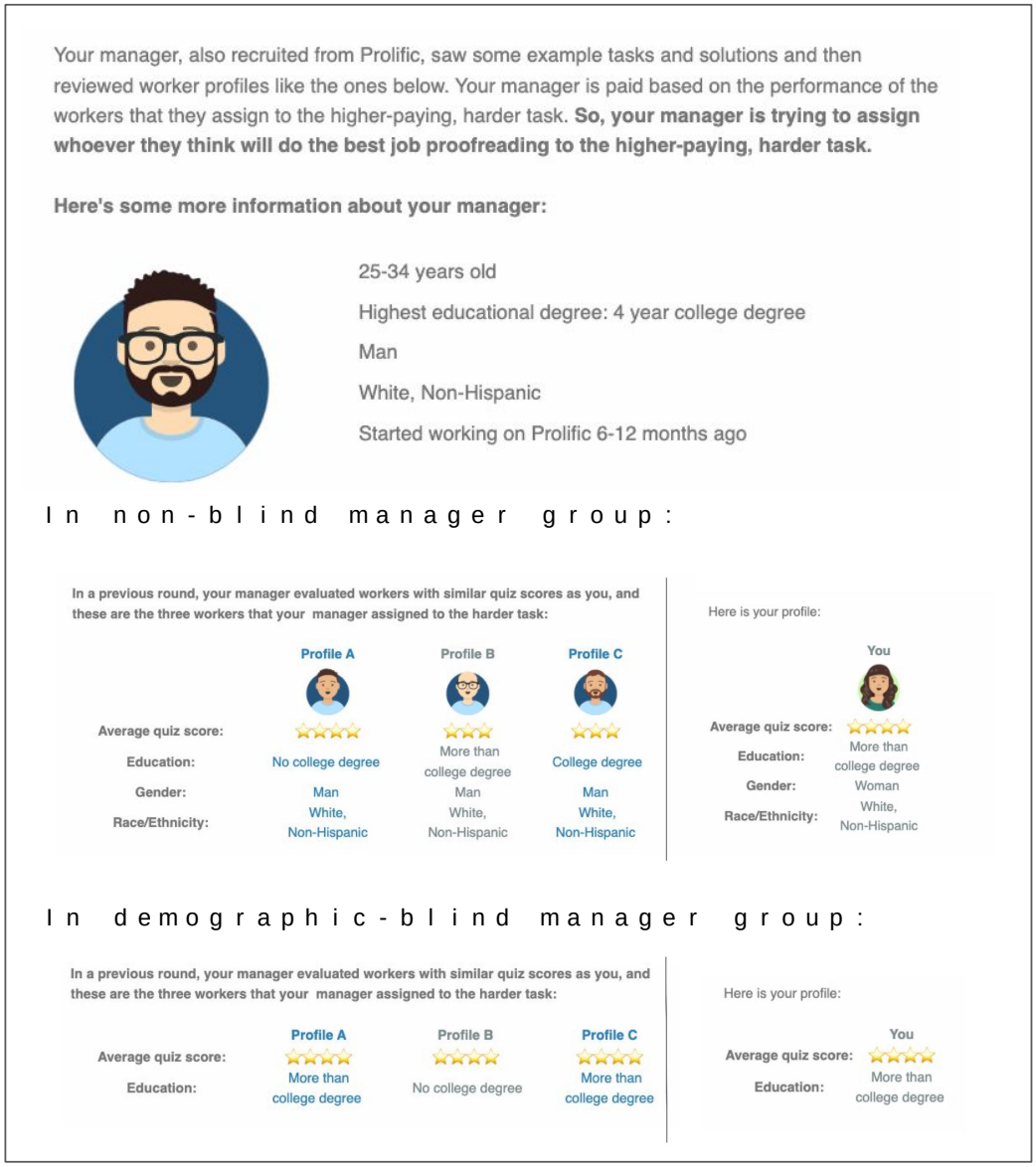
Note: This figure illustrates the treatment arms in the promotion experiment and how they are used to answer various research questions. Randomization is indicated by gray arrows. In the manager sample, workers are randomly assigned to managers. Some previously promoted three white men, some promoted two white men and someone else. This is uncorrelated with other manager characteristics (Appendix Table A14). 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 A5: Timeline of experimental survey (promotion experiment)



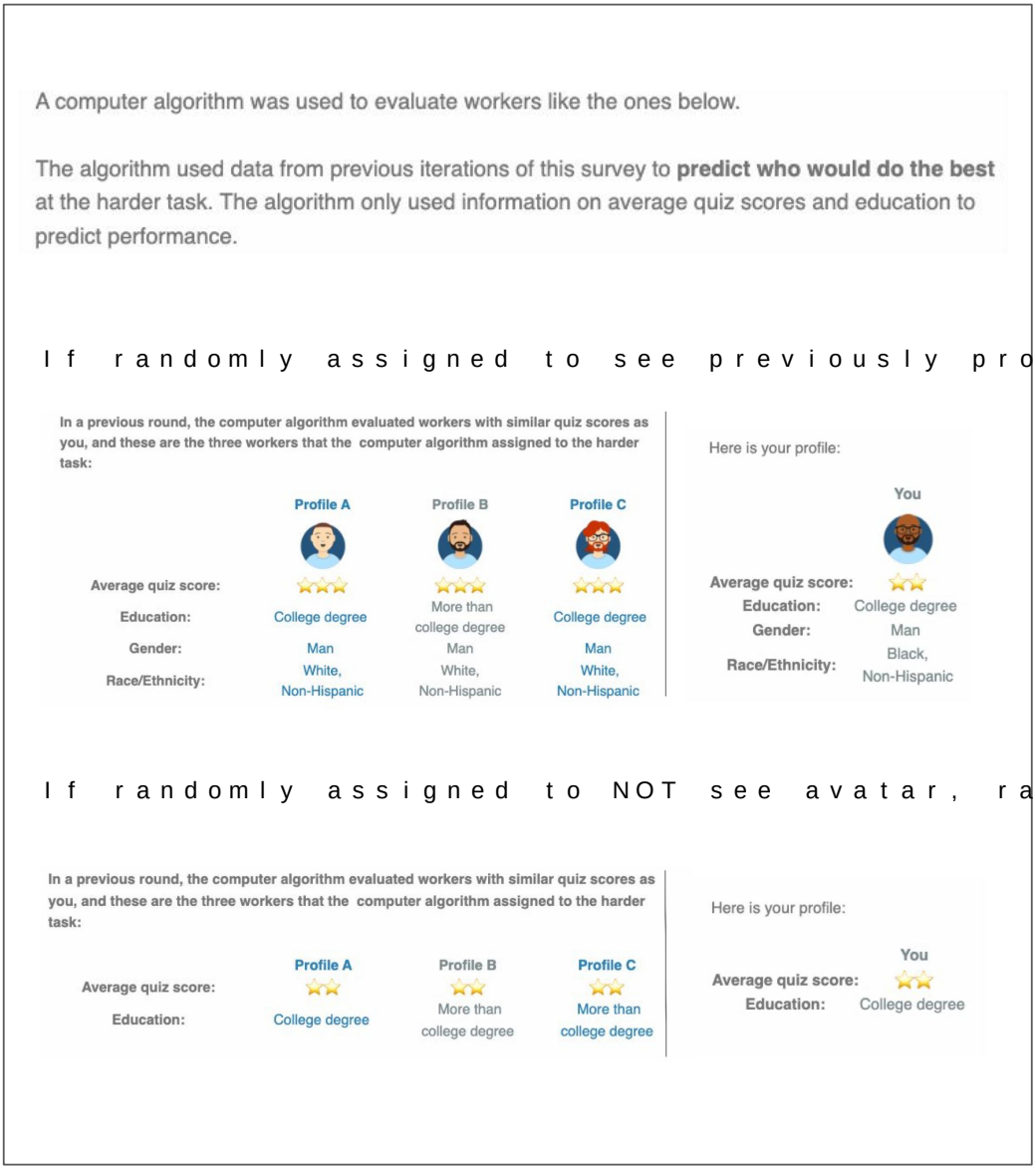
Note: This figure describes the timeline of the experimental survey in the promotion experiment. The experimental survey in the hiring experiment was the same except that workers did not complete the easier proofreading job so I do not observe the corresponding outcome variables, and the survey outcomes additionally included comprehension questions about the hiring procedures.
[Click to return to Section 2 (Research Design)]

Figure A6: Treatment variation in manager groups



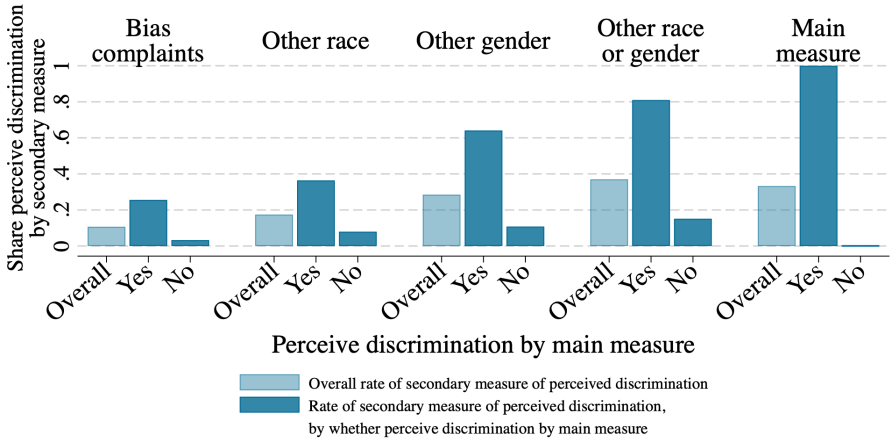
Note: This figure shows what workers saw about their manager, regardless of whether their manager saw avatars, race, and gender or not, and the information they saw about the workers their manager previously promoted in the non-blind manager group and demographic-blind manager group.
[Click to return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

Figure A7: Treatment variation in algorithm groups (promotion experiment)



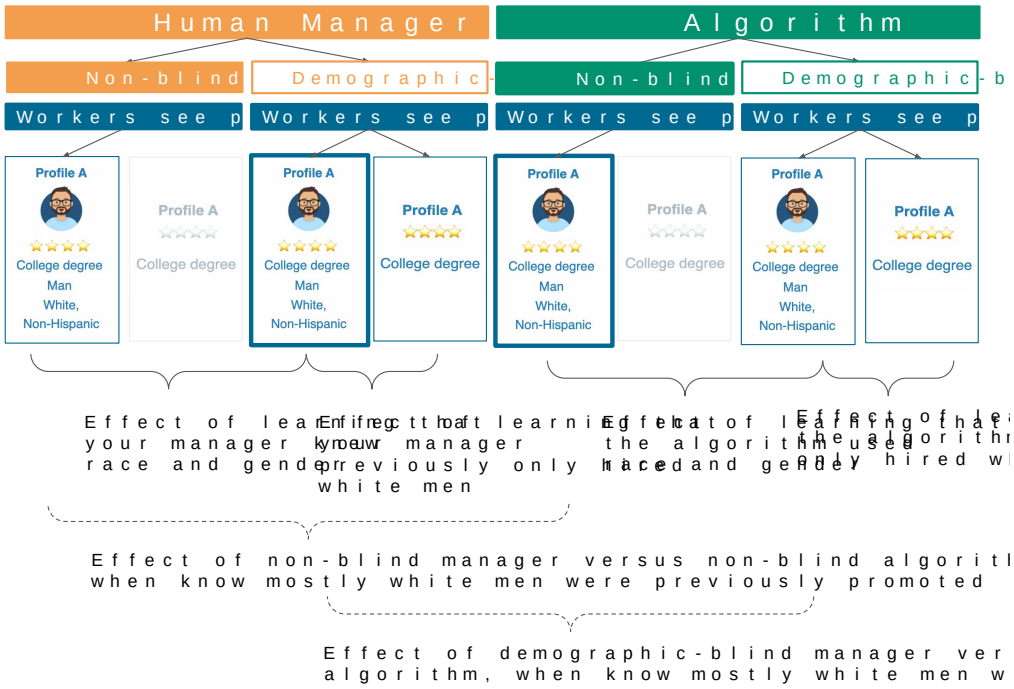
Note: This figure shows what workers saw about the algorithm and the information they saw about the workers their manager previously promoted depending on whether they were randomly assigned to see those workers' avatars, race, and gender or not.
[Click to return to Section 2 (Research Design)]

Figure A8: Correlation of perceived discrimination measures



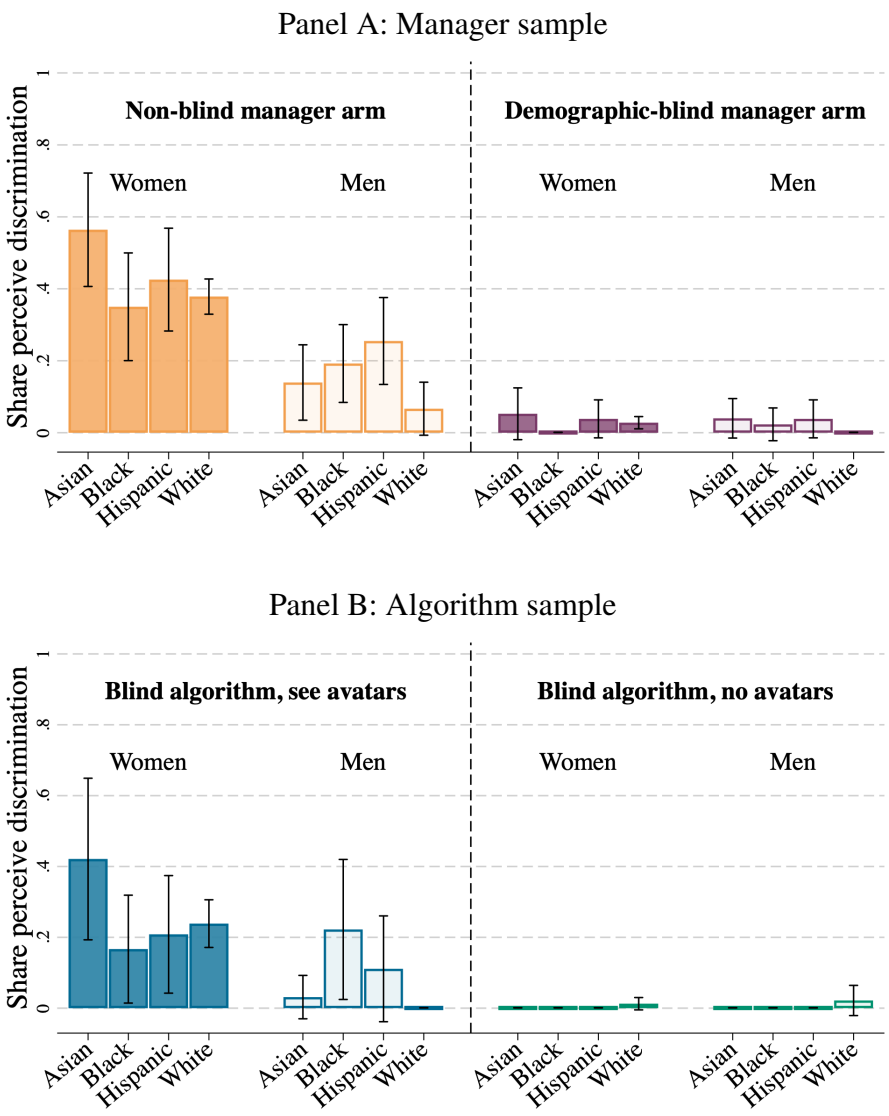
Note: This figure shows the correlation between the primary and secondary measures of perceived discrimination and the overall frequency of each in the non-blind manager arm of the promotion experiment. The primary measure is an indicator for whether a worker mentions demographics in their response to 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 A9: Design of hiring experiment



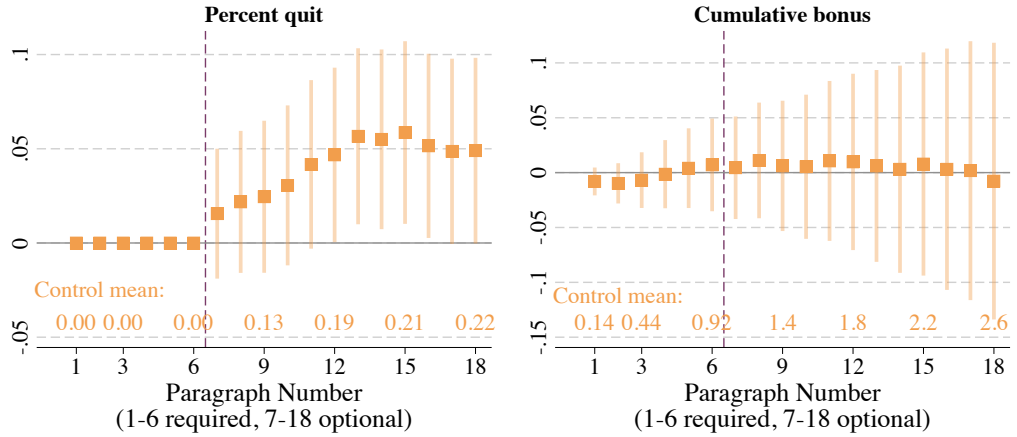
Note: This figure illustrates the treatment arms in the hiring experiment, which replicates the four original treatment arms and adds two new arms, bolded. The grayed-out boxes indicate arms that would be included in the full 2x2x2 factorial design, but are not truthfully or realistically implementable 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 A10: Perceived discrimination by race × gender



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. 95 percent confidence intervals are indicated by the black bars.
[Click to return to Section 3 (Effects of Perceived Discrimination)]

Figure A11: Effects on retention and overall earnings in the promotion experiment



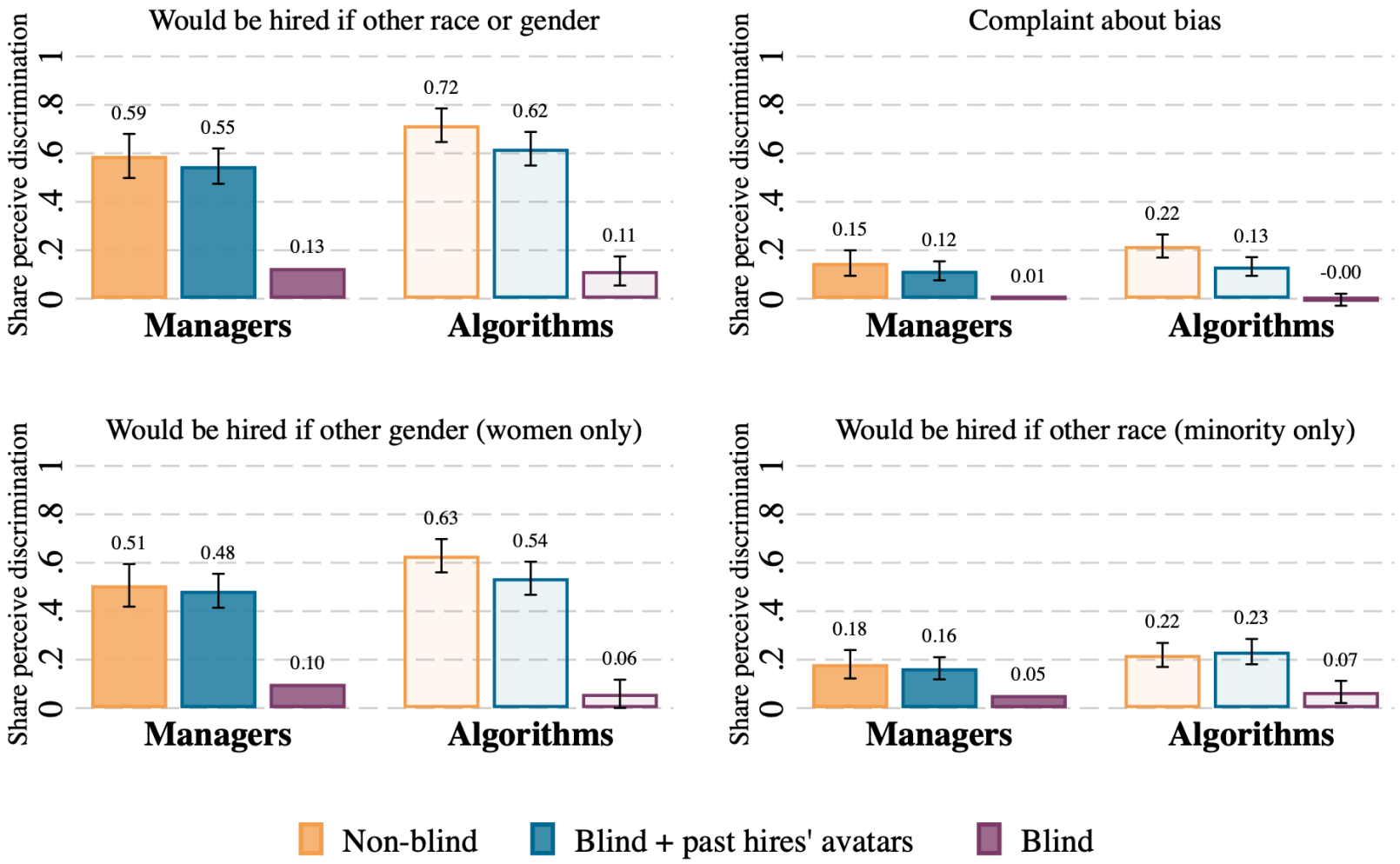
Note: This figure plots the effect of being in the non-blind manager arm relative to the demographic-blind manager arm (Panel A) and the effect of seeing the avatars, race, and gender of previously-promoted workers in the algorithm arm (Panel B) of the promotion experiment on whether workers have quit and how much they have earned, cumulatively by the given paragraph on the x-axis. Workers were required to proofread the first six paragraphs, after which they could quit after any paragraph, so treatment effects absent selection are available in the first six paragraphs. The regressions restrict to workers who would not have been promoted under any promotion procedure. 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. 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3.1 (Effects on Retention and Performance)]

Figure A12: 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. 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. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3.1 (Effects on Retention and Performance)]

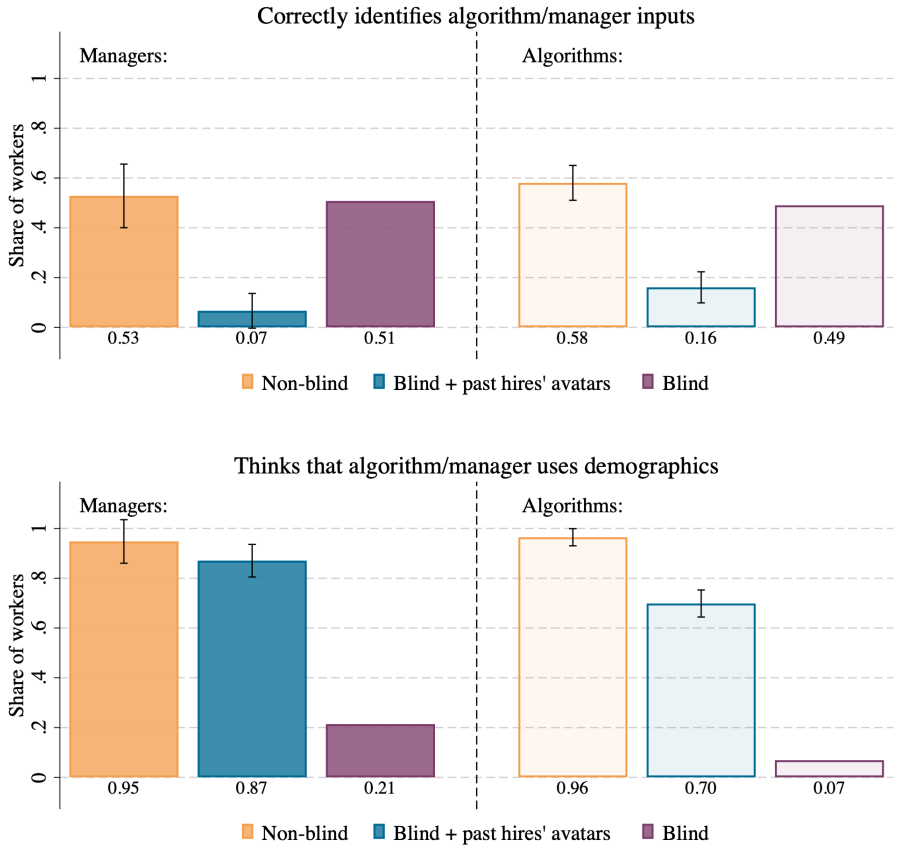
Figure A13: Perceived discrimination in the hiring sample, secondary measures of perceived discrimination



Note: This figure replicates Figure 4 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 A8. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

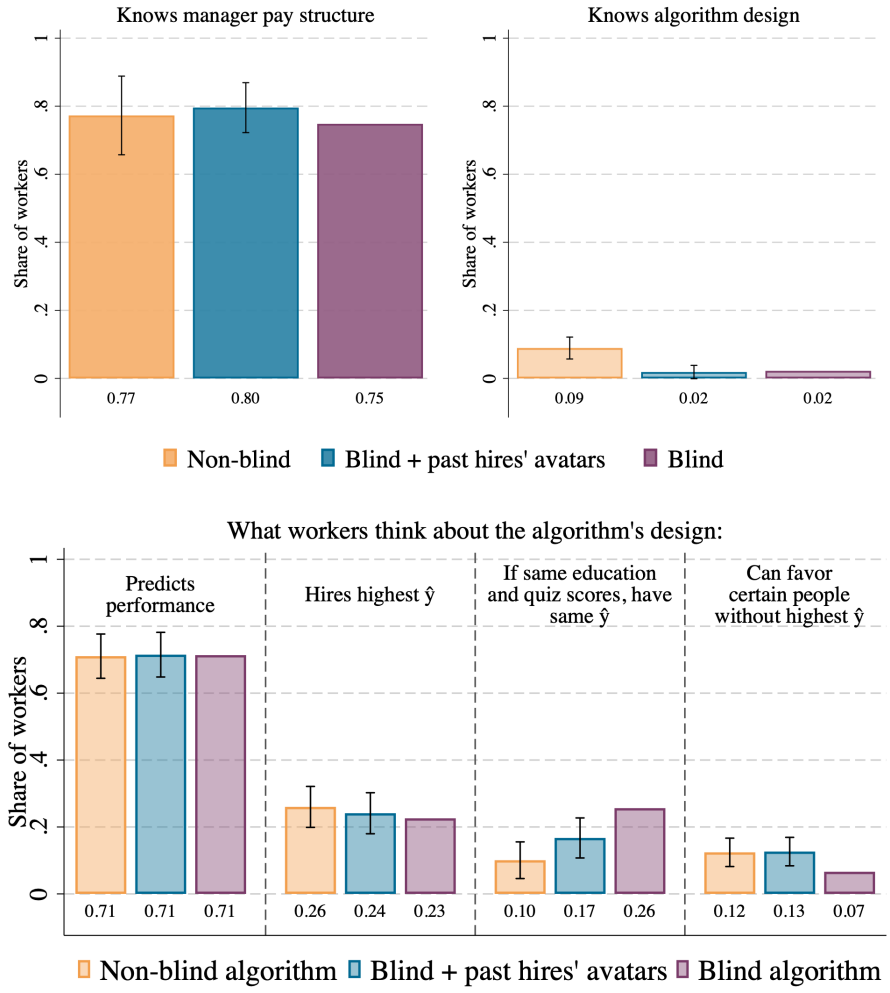
[Click to return to Section 4 (Effects of Hiring Policies)]

Figure A14: 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 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 regressions that 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. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 4 (Effects of Hiring Policies)]

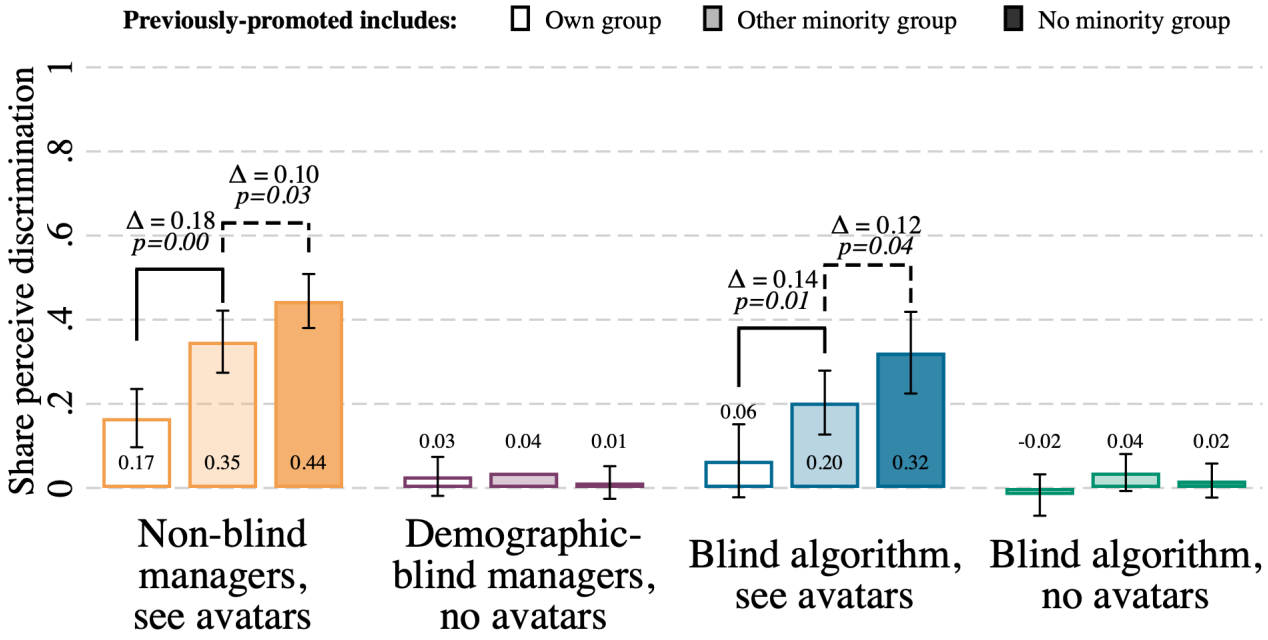
Figure A15: Comprehension of hiring procedure incentives (manager arms) and design (algorithm arms)



Note: This figure shows worker knowledge of managers' incentives and algorithm design. In the top left, it plots the share of workers in the manager arms who know the manager's pay structure. When asked, "what was the basis of your manager's bonus payment?," this requires checking off "The performance in the proofreading task of the workers they hired," and "How many workers they hired showed up to do the proofreading task," and **not** checking off, "How many workers they hired," "Whether they chose workers with the highest screening quiz scores," or "Whether they chose diverse workers." On the right, it plots the share of workers in the algorithm arms who correctly identify the algorithm's design. When asked "what do you know about how the algorithm was designed?," this requires checking off "It was predicting worker performance at the proofreading task," "It hired whoever had the highest predicted [performance]," and, in the demographic-blind decision-maker groups only, "It would provide the same predicted [performance] for any worker with the same education and quiz score," and **not** checking off "It was predicting whether a manager would have hired a worker," or "It could favor certain people and hire them even if they didn't have the highest predicted [performance]." The lower panel plots the share of workers checking off each of these options. Approximately zero workers thought it was predicting managers' previous decisions. 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 regressions that 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. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

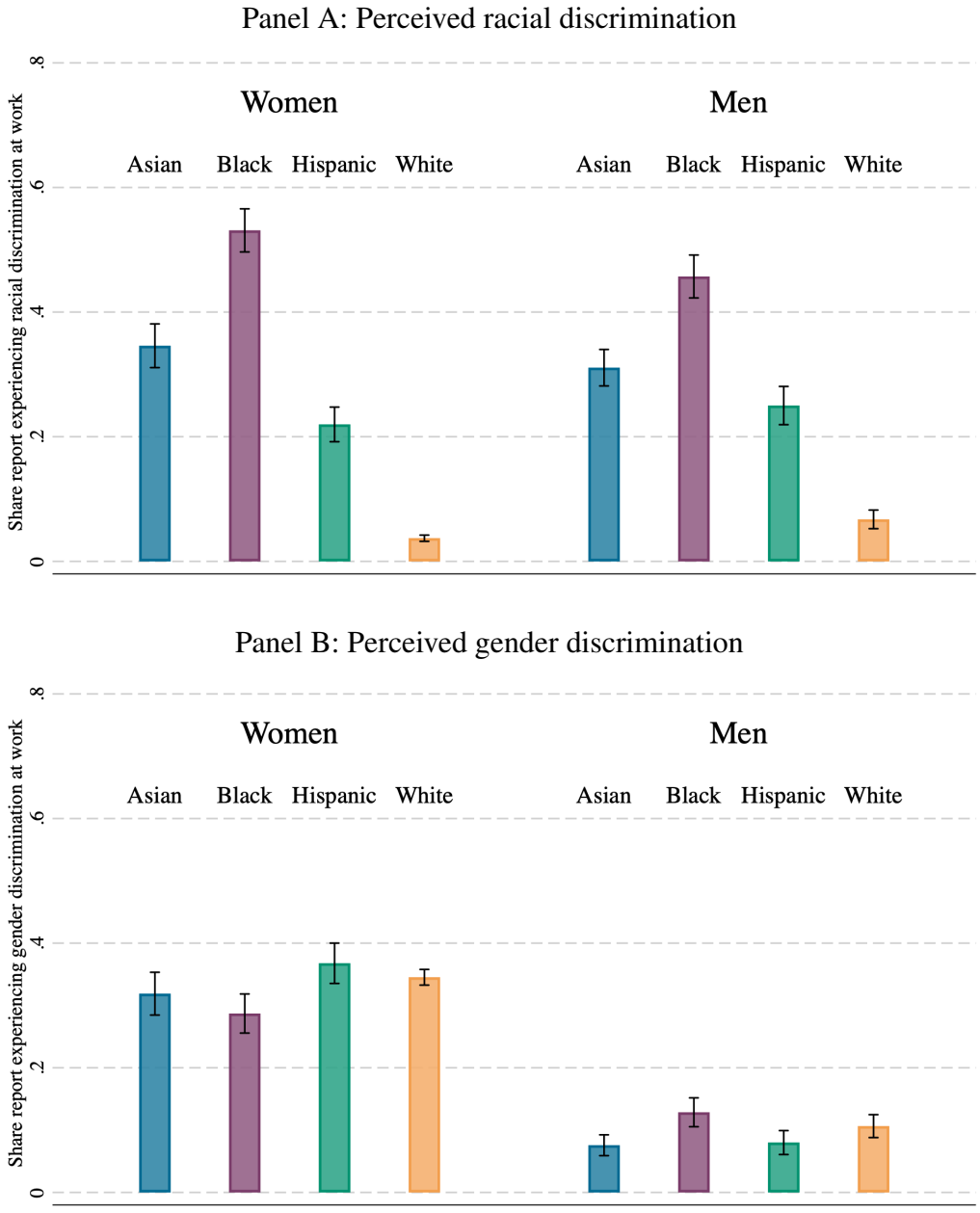
[Click to return to Section 2.4 (Attention and Comprehension)]
[Click to return to Section 4 (Effects of Hiring Policies)]

Figure A16: 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 5, 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. Shares are calculated via regressions that 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. *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 Minority-Group Representation)]

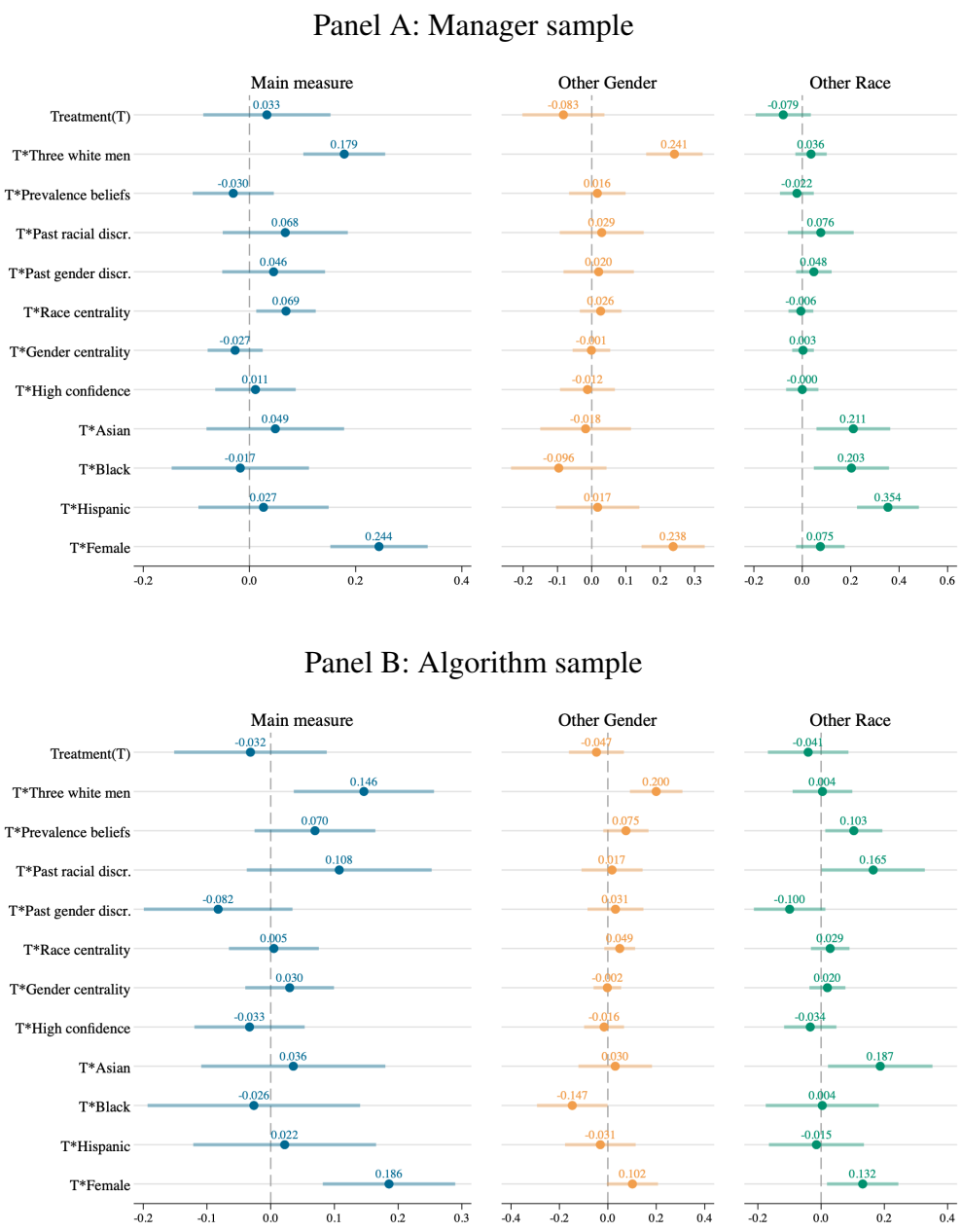
Figure A17: Reported past experiences of discrimination at work in the experimental sample



Note: The sample is 3,240 workers recruited for the promotion experiment on Prolific. The share of each group who reported having experienced discrimination at work in the past is plotted with standard errors in this sample indicated by black bars. Perceived discrimination includes during job search, promotion, termination, and daily work activities. "Asian" refers to those who identify as Asian (no participants identified as Asian and Hispanic), "Black" is those who identify as Black or African American, regardless of whether they identify as Hispanic, "Hispanic" is those who identify as white and Hispanic, and "White" refers to those who identify as white and non-Hispanic. 9 non-binary participants are included with "Women" (those who are more likely to experience and perceive discrimination).

[Click to return to Section 1 (Introduction)]

Figure A18: 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.

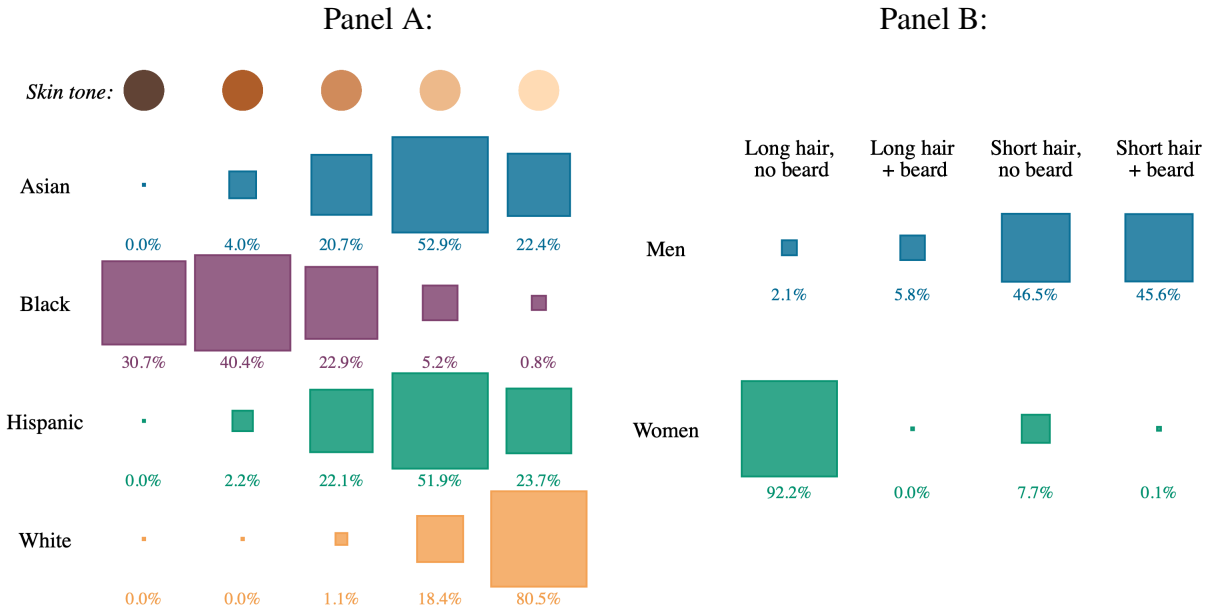
[Click to return to Section 2 (Research Design)]
[Click to return to Section 3 (Effects of Perceived Discrimination)]

Figure A19: Avatar-making procedure

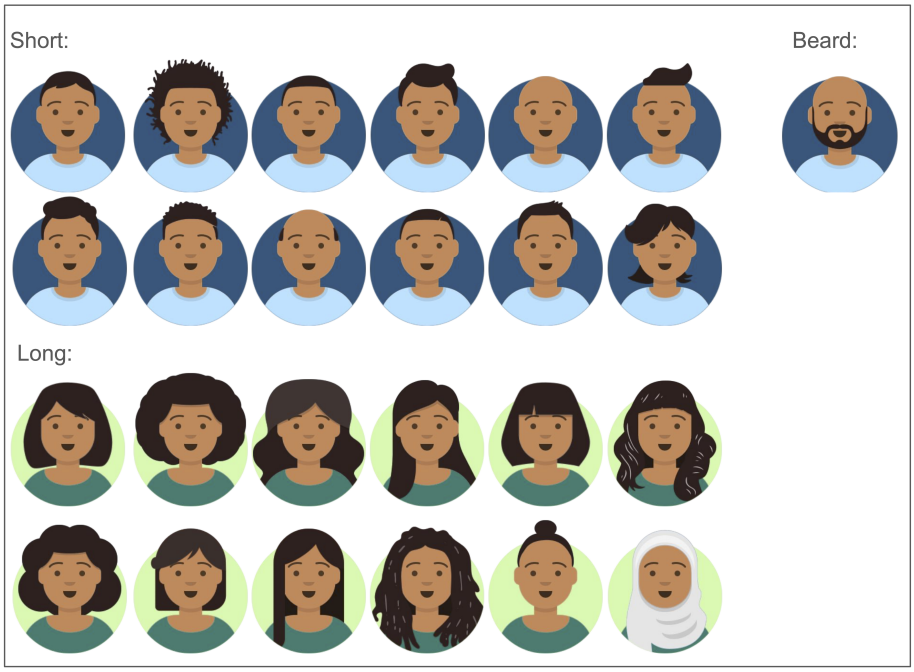


Note: This figure depicts the procedure by which workers created their avatars in the baseline survey.
[Click to return to Section 2 (Research Design)]

Figure A20: Avatar characteristics by self-reported race and gender



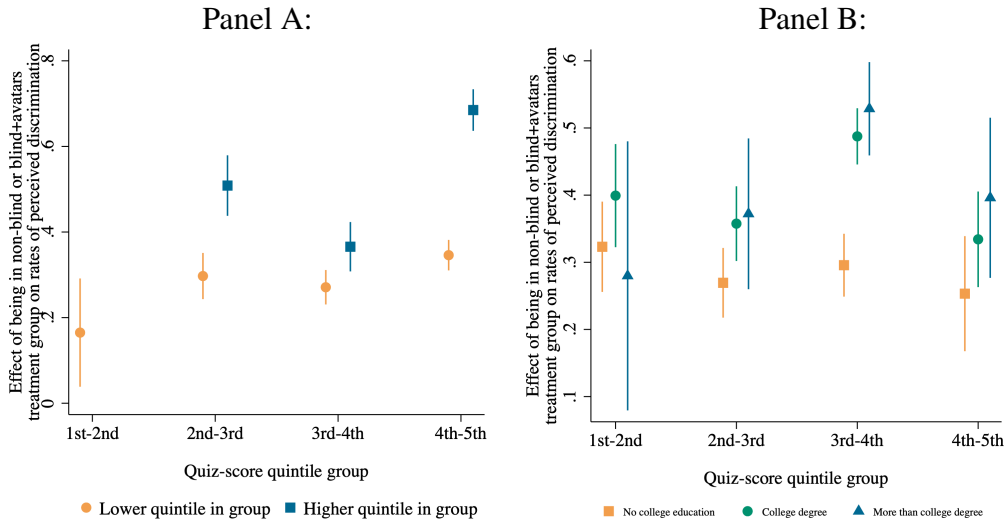
Panel C:



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.

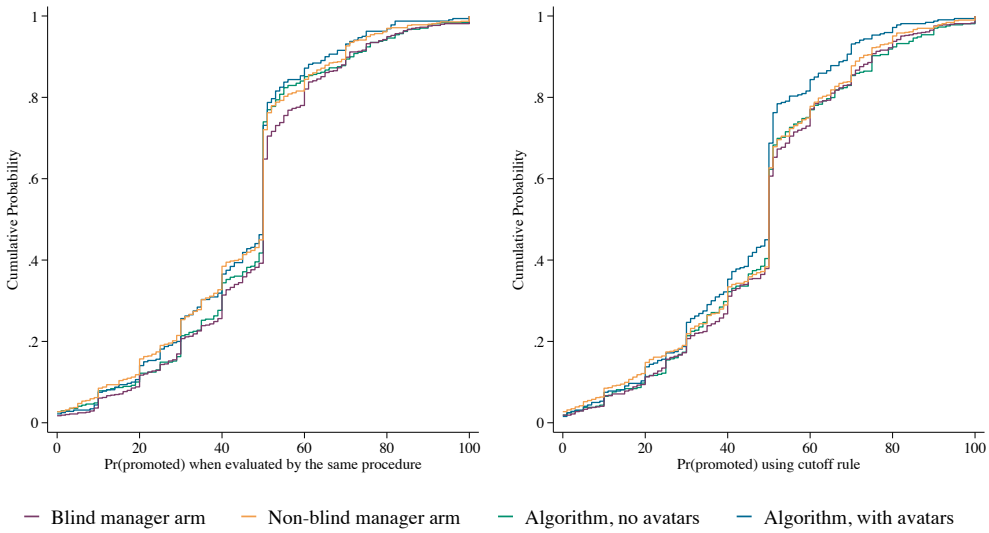
[Click to return to Section 2 (Research Design)]

Figure A21: Perceived discrimination by quiz-score quintile group and quintile



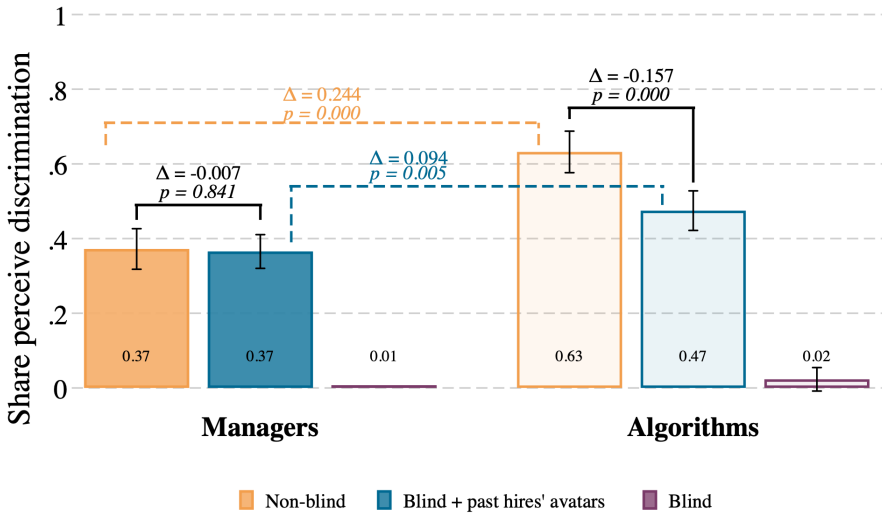
Note: The sample is all workers in the hiring and promotion experiments who would not have been hired/promoted by any of the relevant mechanisms, and the "treatment effect" pools the effect of being in any of the arms with positive perceived discrimination in Figures 1 and 4 relative to the arms with 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 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 A22: CDFs of workers' beliefs about the likelihood of future promotion



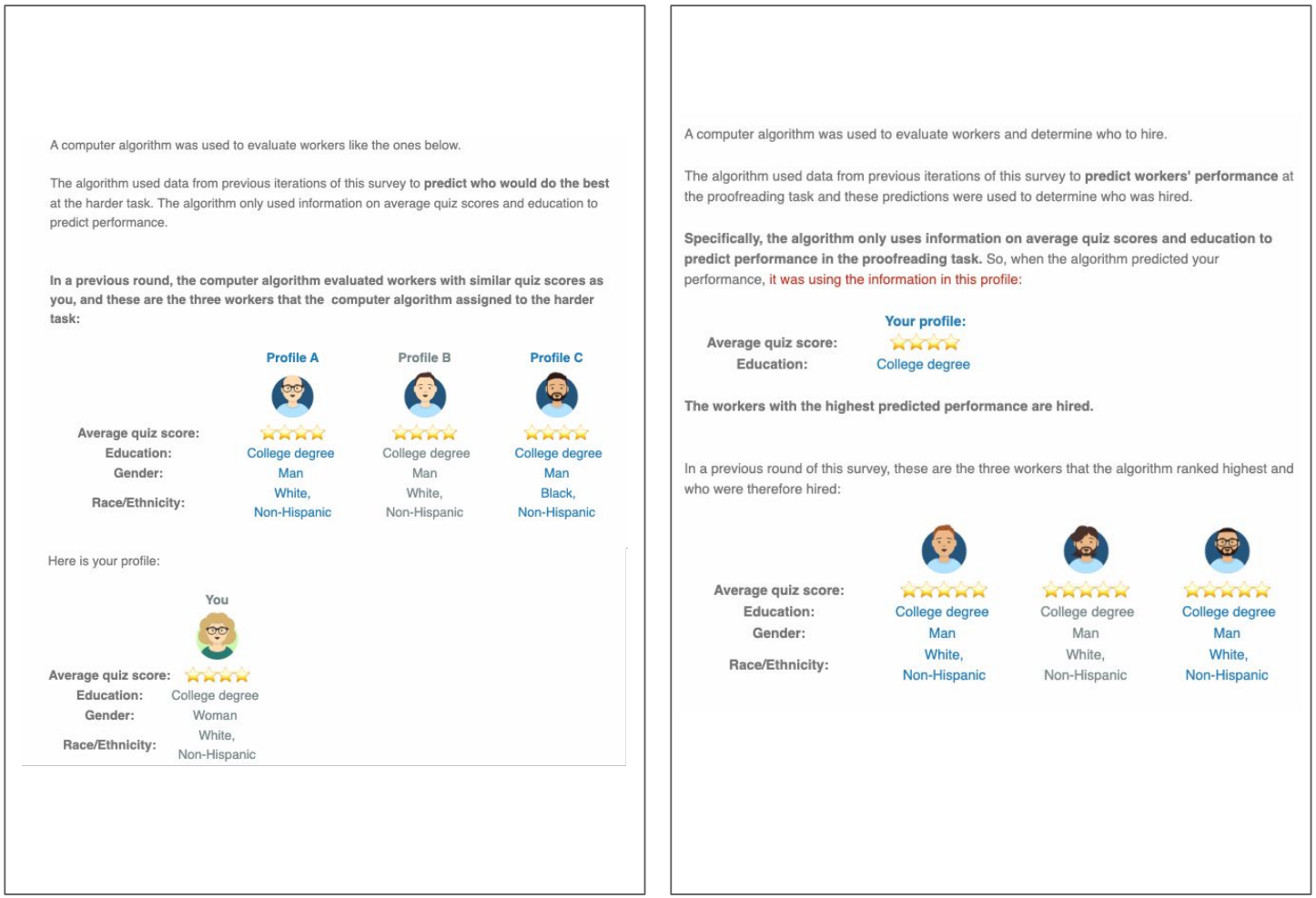
Note: This figure plots the raw CDF of workers' reported beliefs about the likelihood that they will be selected to do the harder task in the future, separately by treatment arm in each experiment. The sample is workers who would not have been selected under any of the relevant promotion or hiring procedures.
[Click to return to Section 3.3.2 (Effects on Future Labor Supply)]

Figure A23: Perceived discrimination in the hiring experiment (full sample)



Note: This figure replicates Figure 4 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 4 (Effects of Hiring Policies)]

Figure A24: Making information about decision-making inputs visually salient in the hiring experiment



Note: This figure shows the only difference in implementation of the replicated treatment arms in the hiring experiment relative to the promotion experiment. In the promotion experiment, workers evaluated by a non-blind algorithm but who saw previously-promoted workers' avatars, race, and gender may have missed the sentence explaining what information the algorithm used (on the left). In the hiring experiment (on the right), this was made equivalently salient by communicating it in words and visually, like the information about the previously-promoted workers, using workers' own profile. This means that workers were also not shown their own demographics in their profile. For consistency, the change in placement of the workers' own profile from the bottom to the top was also changed in the other treatment arms, but in those cases, there was no change in what information was provided.
[Click to return to Section 5 (Effects of Hiring Policies)]

Table A1: 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. 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 A2: 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. *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 A3: Who is selected by each promotion/hiring procedure

Selected by:	Promotion experiment			Hiring experiment			
	Blind manager	Non-blind manager	Blind algorithm	Blind manager	Non-blind manager	Blind algorithm	Non-blind algorithm
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	0.338	0.325	0.274	0.111	0.067	0.156	0.367
Race:							
Asian	0.154	0.201	0.175	0.056	0.089	0.122	0.222
Black	0.107	0.145	0.137	0.178	0.167	0.078	0.178
White, Hispanic	0.098	0.141	0.111	0.033	0.056	0.100	0.356
White, Non-Hispanic	0.641	0.513	0.577	0.733	0.689	0.700	0.244
Quiz score quintile:							
1 star	0.004	0.030	0.000	0.000	0.000	0.000	0.000
2 stars	0.085	0.047	0.077	0.000	0.000	0.000	0.000
3 stars	0.141	0.167	0.154	0.222	0.200	0.200	0.200
4 stars	0.363	0.346	0.346	0.356	0.311	0.300	0.300
5 stars	0.406	0.410	0.423	0.422	0.489	0.500	0.500
Age:							
18-24	0.068	0.103	0.150	0.067	0.089	0.222	0.267
25-34	0.359	0.329	0.380	0.322	0.267	0.300	0.389
35-44	0.269	0.278	0.209	0.278	0.289	0.189	0.156
45-54	0.128	0.158	0.115	0.156	0.222	0.167	0.133
55-64	0.090	0.077	0.064	0.111	0.122	0.100	0.044
65 or older	0.085	0.056	0.081	0.067	0.011	0.022	0.011
N	234	234	234	90	90	90	90

Note: This table shows means of worker characteristics for the workers who are selected by each procedure in each experiment, where the worker characteristics shown are the ones that managers and the algorithm could have used (though only the non-blind procedures could use the information on race and gender).
[Click to return to Section 2 (Research Design)]

Table A4: Balance table, never versus ever promoted/hired

	Promotion experiment			Hiring experiment		
	Main sample	Ever promoted	*p-value*	Main sample	Ever promoted	*p-value*
	(1)	(2)	(1)=(2)	(3)	(4)	(3)=(4)
Male	0.29	0.32	*0.34*	0.13	0.16	*0.30*
Race:						
Asian	0.13	0.19	*0.03*	0.06	0.13	*0.01*
Black	0.14	0.11	*0.12*	0.12	0.07	*0.02*
White, Hispanic	0.14	0.14	*0.74*	0.10	0.17	*0.03*
White, Non-Hispanic	0.59	0.57	*0.60*	0.71	0.63	*0.03*
Married	0.54	0.53	*0.80*	0.58	0.61	*0.39*
Kids	0.43	0.35	*0.02*	0.46	0.42	*0.37*
Education:						
Less than high school	0.01	-0.00	*0.00*	0.01	0.00	*0.00*
High school graduate	0.13	0.02	*0.00*	0.10	0.01	*0.00*
Some college but no degree	0.27	0.05	*0.00*	0.25	0.02	*0.00*
2 year college degree	0.09	0.09	*0.77*	0.10	0.07	*0.26*
4 year college degree	0.37	0.49	*0.00*	0.39	0.45	*0.12*
Professional or Masters degree	0.11	0.29	*0.00*	0.14	0.39	*0.00*
Doctorate	0.01	0.05	*0.00*	0.02	0.05	*0.06*
Income:						
Less than $20,000	0.14	0.09	*0.02*	0.11	0.07	*0.04*
$20,000-$40,000	0.22	0.17	*0.04*	0.18	0.10	*0.01*
$40,000-$60,000	0.18	0.15	*0.26*	0.19	0.24	*0.17*
$60,000-$80,000	0.15	0.12	*0.26*	0.17	0.15	*0.60*
$80,000-$100,000	0.10	0.16	*0.03*	0.13	0.14	*0.70*
$100,000-$120,000	0.07	0.10	*0.19*	0.07	0.11	*0.13*
$120,000-$140,000	0.04	0.05	*0.38*	0.06	0.03	*0.02*
$140,000-$160,000	0.04	0.08	*0.03*	0.04	0.07	*0.12*
More than $160,000	0.06	0.08	*0.20*	0.07	0.10	*0.20*
Age:						
18-24	0.17	0.11	*0.01*	0.15	0.16	*0.82*
25-34	0.32	0.38	*0.07*	0.33	0.29	*0.31*
35-44	0.22	0.24	*0.38*	0.22	0.24	*0.57*
45-54	0.14	0.12	*0.37*	0.17	0.17	*0.91*
55-64	0.11	0.08	*0.17*	0.10	0.09	*0.65*
65 or older	0.05	0.07	*0.28*	0.04	0.06	*0.33*
Employment:						
Currently employed outside Prolific	0.67	0.76	*0.00*	0.69	0.81	*0.00*
Ever employed outside Prolific	0.96	0.98	*0.05*	0.97	0.99	*0.01*
Past experienced discrimination:						
Job search	0.25	0.29	*0.22*	–	–	–
Promotion	0.18	0.17	*0.78*	–	–	–
Termination	0.10	0.07	*0.14*	–	–	–
Daily work activities	0.33	0.38	*0.12*	–	–	–
Any discrimination	0.50	0.56	*0.14*	–	–	–
Beliefs about prevalence (own group)	0.46	0.44	*0.54*	–	–	–
N	2080	237		2527	153	

Note: This table shows means of worker demographic characteristics, all measured in the baseline survey, for workers who are included in the analysis sample (because they would not have been selected by any of the relevant promotion or hiring procedures) and those who are not, separately for the promotion and hiring experiments. The *p*-values test the null that ever-selected and never-selected workers are the same and are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 2 (Research Design)]

Table A5: 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,	0.182***	0.195***	0.168***	0.081*
see avatars	(0.016)	(0.021)	(0.033)	(0.042)
Control mean	0.010	0.014	0.000	0.000
Panel B: Yes, would be promoted if different race or gender				
Non-blind manager,	0.251***	0.288***	0.166***	0.084
see avatars	(0.024)	(0.029)	(0.056)	(0.052)
Control mean	0.127	0.130	0.159	0.000
Panel C: Complaint about job assignment mentions discrimination				
Non-blind manager,	0.099***	0.116***	0.050**	0.059
see avatars	(0.013)	(0.016)	(0.022)	(0.049)
Control mean	0.003	0.004	0.000	0.000
*Panel D: Members of my race * gender group underrepresented among promoted*				
Non-blind manager,	0.162***	0.208***	-0.009	0.015
see avatars	(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,	0.157***	0.196***	0.024	0.327**
see avatars	(0.026)	(0.030)	(0.062)	(0.141)
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 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 1. 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 (Effects of Perceived Discrimination)]

Table A6: Effects on retention, gender and racial heterogeneity

	Num paragraphs	More than 6	Did all 18
	(1)	(2)	(3)
Panel A: Race and gender heterogeneity			
Non-blind manager arm*White women	-0.645**	-0.039*	-0.046
	(0.316)	(0.021)	(0.032)
Non-blind manager arm*Minority women	-1.001*	-0.029	-0.120**
	(0.557)	(0.038)	(0.057)
Non-blind manager arm*Minority men	0.808	0.065	0.065
	(0.560)	(0.042)	(0.054)
p-values:			
White=minority, women	*0.566*	*0.809*	*0.247*
Men=women, minority	*0.020*	*0.094*	*0.016*
N	1293	1293	1293
Panel B: Gender heterogeneity			
Non-blind manager arm*Women	-0.741***	-0.036*	-0.066**
	(0.285)	(0.019)	(0.029)
Non-blind manager arm*Minority men	0.808	0.065	0.065
	(0.560)	(0.042)	(0.054)
p-values:			
Men=women	*0.011*	*0.026*	*0.027*
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 3.3.1 (Effects on Retention and Performance)]

Table A7: Effects on effort and performance in the manager sample of the promotion experiment, gender and racial heterogeneity

	Minutes	Clicks	Correct	Incorrect	Bonus
	(1)	(2)	(3)	(4)	(5)
Panel A: Race and gender heterogeneity					
Non-blind manager arm*White women	-0.060	1.107	-0.055	0.070	-0.017
	(0.073)	(2.390)	(0.296)	(0.257)	(0.028)
Non-blind manager arm*Minority women	-0.012	4.397	0.803	0.798**	0.043
	(0.116)	(4.401)	(0.522)	(0.406)	(0.049)
Non-blind manager arm*Minority men	0.099	-4.937*	1.320***	-0.109	0.080*
	(0.119)	(2.622)	(0.469)	(0.487)	(0.045)
p-values:					
White=minority, women	*0.721*	*0.513*	*0.139*	*0.109*	*0.268*
Men=women, minority	*0.496*	*0.061*	*0.454*	*0.154*	*0.566*
N	1295	1295	1295	1295	1295
Panel B: Gender heterogeneity					
Non-blind manager arm*Women	-0.048	1.989	0.175	0.265	-0.001
	(0.063)	(2.098)	(0.267)	(0.229)	(0.025)
Non-blind manager arm*Minority men	0.099	-4.933*	1.321***	-0.108	0.080*
	(0.119)	(2.619)	(0.469)	(0.487)	(0.045)
p-values:					
Men=women	*0.259*	*0.033*	*0.032*	*0.485*	*0.106*
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 A6. 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 3.3.1 (Effects on Retention and Performance)]

Table A8: 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	-0.079***	-0.053*	-0.047*	-0.063**	-0.026
	(0.028)	(0.029)	(0.029)	(0.029)	(0.027)
N	1380	1380	1380	1380	1389
Control mean	0.657	0.561	0.584	0.532	0.689

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 panel A) and the effect of seeing that previously-promoted workers were mostly white men in the algorithm arm (in panel B) 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 levels indicated by *, **, and ***, respectively.

[Click to return to Section 3.3.2 (Effects on Future Labor Supply)]

Table A9: 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	0.27	0.29	*0.571*	0.31	0.23	*0.015*	0.30	0.32	*0.767*	0.36	0.33	*0.580*
Race:												
Asian	0.13	0.11	*0.391*	0.15	0.09	*0.014*	0.18	0.12	*0.137*	0.14	0.14	*0.941*
Black	0.15	0.12	*0.241*	0.14	0.13	*0.918*	0.12	0.16	*0.238*	0.17	0.16	*0.727*
White, Hispanic	0.14	0.14	*0.997*	0.14	0.16	*0.626*	0.14	0.11	*0.350*	0.15	0.14	*0.863*
White, Non-Hispanic	0.59	0.64	*0.168*	0.57	0.62	*0.170*	0.56	0.62	*0.351*	0.55	0.56	*0.741*
Married	0.51	0.52	*0.736*	0.55	0.53	*0.536*	0.53	0.57	*0.508*	0.60	0.50	*0.066*
Kids	0.49	0.40	*0.024*	0.43	0.42	*0.706*	0.40	0.48	*0.162*	0.44	0.38	*0.277*
Education:												
Less than high school	0.01	0.01	*0.791*	0.01	0.01	*0.409*	0.02	-0.00	*0.164*	0.01	0.01	*0.974*
High school graduate	0.16	0.13	*0.248*	0.14	0.14	*0.958*	0.12	0.15	*0.408*	0.11	0.11	*0.924*
Some college but no degree	0.26	0.27	*0.852*	0.28	0.29	*0.782*	0.25	0.32	*0.175*	0.23	0.27	*0.386*
2 year college degree	0.09	0.10	*0.563*	0.10	0.10	*0.911*	0.07	0.07	*0.827*	0.10	0.09	*0.557*
4 year college degree	0.34	0.34	*0.903*	0.36	0.38	*0.536*	0.42	0.38	*0.533*	0.41	0.36	*0.279*
Professional or Masters degree	0.12	0.13	*0.551*	0.10	0.08	*0.318*	0.11	0.08	*0.404*	0.12	0.15	*0.341*
Doctorate	0.02	0.01	*0.523*	0.01	0.01	*0.970*	0.01	0.00	*0.229*	0.01	0.01	*0.704*
Income:												
Less than $20,000	0.13	0.16	*0.375*	0.13	0.13	*0.867*	0.14	0.14	*0.898*	0.11	0.21	*0.010*
$20,000-$40,000	0.21	0.20	*0.771*	0.21	0.29	*0.027*	0.25	0.18	*0.154*	0.23	0.17	*0.121*
$40,000-$60,000	0.20	0.19	*0.785*	0.17	0.16	*0.700*	0.20	0.18	*0.659*	0.17	0.14	*0.450*
$60,000-$80,000	0.14	0.16	*0.419*	0.15	0.14	*0.514*	0.15	0.17	*0.567*	0.15	0.11	*0.250*
$80,000-$100,000	0.09	0.12	*0.265*	0.14	0.08	*0.042*	0.07	0.11	*0.341*	0.09	0.11	*0.495*
$100,000-$120,000	0.09	0.05	*0.031*	0.07	0.08	*0.855*	0.05	0.07	*0.434*	0.07	0.08	*0.682*
$120,000-$140,000	0.03	0.02	*0.235*	0.04	0.04	*0.887*	0.04	0.03	*0.573*	0.05	0.07	*0.508*
$140,000-$160,000	0.04	0.05	*0.308*	0.03	0.04	*0.423*	0.03	0.04	*0.608*	0.05	0.03	*0.192*
More than $160,000	0.06	0.05	*0.414*	0.05	0.04	*0.542*	0.06	0.08	*0.690*	0.07	0.09	*0.489*
Age:												
18-24	0.14	0.18	*0.173*	0.17	0.17	*0.976*	0.20	0.14	*0.234*	0.19	0.15	*0.395*
25-34	0.30	0.35	*0.135*	0.32	0.29	*0.398*	0.29	0.40	*0.045*	0.30	0.34	*0.501*
35-44	0.24	0.18	*0.057*	0.22	0.22	*0.963*	0.26	0.16	*0.046*	0.19	0.22	*0.516*
45-54	0.14	0.13	*0.617*	0.15	0.14	*0.514*	0.13	0.10	*0.388*	0.14	0.16	*0.632*
55-64	0.11	0.11	*0.830*	0.09	0.14	*0.044*	0.06	0.14	*0.007*	0.14	0.09	*0.202*
65 or older	0.06	0.05	*0.441*	0.04	0.04	*0.887*	0.06	0.05	*0.661*	0.04	0.04	*0.955*
Employment:												
Currently employed outside Prolific	0.64	0.65	*0.817*	0.66	0.65	*0.856*	0.69	0.77	*0.141*	0.71	0.62	*0.081*
Ever employed outside Prolific	0.93	0.97	*0.020*	0.96	0.96	*0.815*	0.97	0.99	*0.223*	0.95	0.95	*0.968*
Satisfied with current employer	0.78	0.80	*0.501*	0.78	0.77	*0.770*	0.72	0.80	*0.211*	0.78	0.76	*0.714*
Satisfied with most recent employer	0.59	0.65	*0.374*	0.64	0.68	*0.595*	0.63	0.70	*0.599*	0.67	0.62	*0.623*
N	376	319		422	272		216	104		220	151	

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). *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 2 (Research Design)]

Table A10: 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,	-0.109***	-0.116***	-0.146**	0.099
see avatars	(0.024)	(0.028)	(0.057)	(0.104)
Control mean	0.414	0.410	0.428	0.408
Panel B: Free response, first answer is higher quiz score				
Non-blind manager,	-0.056**	-0.079**	0.005	-0.041
see avatars	(0.027)	(0.031)	(0.066)	(0.110)
Control mean	0.484	0.488	0.476	0.469
Panel C: Free response, first answer is it was random				
Non-blind manager,	0.000	0.007	0.001	-0.081
see avatars	(0.011)	(0.013)	(0.021)	(0.064)
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. 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 1. 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 (Effects of Perceived Discrimination)]

Table A11: 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	0.006	-0.013	-0.038
	(0.035)	(0.040)	(0.029)
N	1389	1382	1389
Control mean	0.011	-0.013	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 panel A) and the effect of seeing that previously-promoted workers were mostly white men in the algorithm arm (in panel B) 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 levels indicated by *, **, and ***, respectively.
[Click to return to Section 3.3.1 (Effects on Retention and Performance)]

Table A12: 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 arm*White women	-0.004	0.006	-0.015
	(0.048)	(0.052)	(0.039)
Non-blind manager arm*Minority women	-0.064	-0.055	-0.056
	(0.079)	(0.094)	(0.062)
Non-blind manager arm*Minority men	0.106	-0.026	-0.062
	(0.071)	(0.085)	(0.059)
p-values:			
White=minority, women	*0.506*	*0.563*	*0.562*
Men=women, minority	*0.110*	*0.818*	*0.937*
N	1295	1288	1295
Panel B: Gender heterogeneity			
Non-blind manager arm*Women	-0.020	-0.010	-0.026
	(0.042)	(0.047)	(0.034)
Non-blind manager arm*Minority men	0.106	-0.026	-0.062
	(0.071)	(0.085)	(0.059)
p-values:			
Men=women	*0.124*	*0.870*	*0.580*
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 A11, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A6. 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 3.3.1 (Effects on Retention and Performance)]

Table A13: 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 arm*White women	0.033**	0.012**	0.025	0.051*	0.038	-0.234
	(0.014)	(0.006)	(0.019)	(0.027)	(0.039)	(0.317)
Non-blind manager arm*Minority women	-0.007	0.011	-0.022	0.115***	0.009	-0.585
	(0.021)	(0.011)	(0.029)	(0.044)	(0.062)	(0.442)
Non-blind manager arm*Minority men	0.025	0.013	-0.014	-0.015	0.103*	-0.001
	(0.023)	(0.012)	(0.032)	(0.046)	(0.058)	(0.496)
p-values:						
White=minority, women	*0.108*	*0.912*	*0.182*	*0.201*	*0.680*	*0.500*
Men=women, minority	*0.298*	*0.865*	*0.853*	*0.043*	*0.261*	*0.373*
N	1247	1235	1256	957	1289	1289
Panel B: Gender heterogeneity						
Non-blind manager arm*Women	0.022*	0.012**	0.012	0.067***	0.030	-0.329
	(0.012)	(0.005)	(0.017)	(0.023)	(0.034)	(0.269)
Non-blind manager arm*Minority men	0.025	0.013	-0.014	-0.015	0.103*	-0.002
	(0.023)	(0.012)	(0.032)	(0.046)	(0.058)	(0.496)
p-values:						
Men=women	*0.902*	*0.890*	*0.461*	*0.109*	*0.268*	*0.549*
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 A6. 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 3.3.2 (Effects on Future Labor Supply)]

Table A14: Correlations of manager characteristics with previously selected workers' characteristics

	3 white men	3 men	3 white	3 max stars	2 max stars	1 max stars	3 grad sch	3 college	Any grad sch	Any coll	Any no coll
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age: 25-34	-0.026	-0.081	0.123	0.086	-0.080	-0.005	-0.049	-0.169	0.198	-0.062	0.131
	(0.184)	(0.195)	(0.175)	(0.159)	(0.150)	(0.064)	(0.130)	(0.124)	(0.152)	(0.161)	(0.163)
Age: 35-44	0.415**	0.201	0.394**	0.083	-0.119	0.035	0.183	0.007	0.005	-0.122	-0.088
	(0.194)	(0.205)	(0.185)	(0.168)	(0.158)	(0.067)	(0.137)	(0.131)	(0.160)	(0.169)	(0.172)
Age: 45-54	-0.018	-0.022	0.142	0.083	-0.052	-0.031	-0.043	-0.122	0.096	0.036	-0.001
	(0.230)	(0.243)	(0.218)	(0.198)	(0.187)	(0.080)	(0.162)	(0.155)	(0.189)	(0.200)	(0.204)
Age: 55-64	0.237	-0.073	0.044	-0.074	0.065	0.009	0.507**	-0.188	0.237	-0.419	-0.202
	(0.291)	(0.307)	(0.276)	(0.251)	(0.237)	(0.101)	(0.205)	(0.196)	(0.239)	(0.253)	(0.258)
Age: 65 or older	-0.212	0.512	-0.419	0.340	-0.336	-0.004	-0.085	-0.148	-0.796*	0.201	0.790*
	(0.523)	(0.552)	(0.496)	(0.451)	(0.426)	(0.181)	(0.369)	(0.352)	(0.430)	(0.455)	(0.463)
Educ: 4 year college degree	-0.029	-0.041	0.148	-0.299	0.299	0.000	-0.004	-0.159	0.094	-0.095	0.170
	(0.252)	(0.266)	(0.239)	(0.217)	(0.205)	(0.087)	(0.178)	(0.169)	(0.207)	(0.219)	(0.223)
Educ: Doctorate	-0.479	0.545	-0.315	0.115	-0.102	-0.013	-0.597	-0.119	0.061	0.525	0.161
	(0.561)	(0.592)	(0.533)	(0.484)	(0.457)	(0.194)	(0.396)	(0.378)	(0.462)	(0.489)	(0.497)
Educ: High school graduate	0.179	-0.086	0.622**	-0.110	0.120	-0.010	0.304	-0.105	-0.013	-0.277	0.167
	(0.284)	(0.300)	(0.270)	(0.245)	(0.231)	(0.098)	(0.200)	(0.191)	(0.234)	(0.247)	(0.252)
Educ: Professional or Masters degree	0.006	0.051	0.321	-0.128	0.165	-0.037	-0.146	-0.054	-0.070	0.176	0.026
	(0.286)	(0.302)	(0.271)	(0.246)	(0.233)	(0.099)	(0.202)	(0.192)	(0.235)	(0.249)	(0.253)
Educ: Some college but no degree	-0.069	-0.106	0.347	-0.139	0.155	-0.015	-0.184	-0.154	0.165	0.020	0.360
	(0.270)	(0.285)	(0.257)	(0.233)	(0.220)	(0.094)	(0.191)	(0.182)	(0.222)	(0.235)	(0.240)
On Prolific: 3-6 months ago	-0.153	-0.302	0.157	0.125	-0.212	0.088	0.020	0.338**	-0.501***	0.224	-0.061
	(0.209)	(0.221)	(0.199)	(0.180)	(0.170)	(0.073)	(0.148)	(0.141)	(0.172)	(0.182)	(0.186)
On Prolific: 6-12 months ago	0.515*	0.210	0.419	-0.314	0.057	0.257**	0.243	0.298	-0.185	-0.062	-0.057
	(0.286)	(0.302)	(0.271)	(0.246)	(0.233)	(0.099)	(0.202)	(0.192)	(0.235)	(0.249)	(0.253)
On Prolific: more than 2 years ago	-0.177	-0.274	0.076	-0.026	0.026	-0.000	0.053	0.109	-0.056	0.002	-0.090
	(0.175)	(0.185)	(0.166)	(0.151)	(0.143)	(0.061)	(0.124)	(0.118)	(0.144)	(0.153)	(0.155)
On Prolific: within the last 3 months	-0.124	-0.119	0.056	-0.086	0.085	0.001	0.226*	0.057	-0.013	-0.059	-0.220
	(0.185)	(0.195)	(0.176)	(0.160)	(0.151)	(0.064)	(0.131)	(0.125)	(0.152)	(0.161)	(0.164)
N	75	75	75	75	75	75	75	75	75	75	75
F-statistic p-value	0.141	0.674	0.151	0.867	0.892	0.592	0.021	0.738	0.205	0.475	0.433
R-squared	0.259	0.156	0.255	0.120	0.114	0.169	0.333	0.145	0.240	0.188	0.195

Note: This table shows estimated coefficients from regressions of the characteristics of workers each manager promoted in the historical sample on the manager characteristics that were observable to workers (other than race and gender, which were constant). The dependent variables are indicators for whether the three workers are all white men, all men, and all white in the first three columns, and the rest are the previously promoted worker educational and performance measures that are included as controls in the main specification. 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 2 (Research Design)]

Table A15: 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	0.086***	0.107***	0.044	–
	(0.019)	(0.025)	(0.043)	
Control mean	0.003	0.004	0.000	0.000
Panel B: Yes, would be promoted if different race or gender				
See avatars	0.166***	0.204***	0.111	0.003
	(0.030)	(0.036)	(0.079)	(0.066)
Control mean	0.054	0.045	0.096	0.022
Panel C: Complaint about job assignment mentions discrimination				
See avatars	0.039***	0.042***	0.030	–
	(0.012)	(0.015)	(0.032)	
Control mean	0.000	0.000	0.000	0.000
*Panel D: Members of my race * gender group underrepresented among promoted*				
See avatars	0.107***	0.115**	0.114	0.041
	(0.039)	(0.049)	(0.107)	(0.110)
Control mean	0.294	0.285	0.446	0.065
Panel E: White men are overrepresented among promoted				
See avatars	0.127***	0.134**	0.258**	-0.273
	(0.043)	(0.053)	(0.114)	(0.202)
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 A5. 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 levels indicated by *, **, and ***, respectively.
[Click to return to Section 6 (Effects of Perceived Algorithmic Discrimination)]

Table A16: Effects on other reasons for non-promotion in algorithm sample of the promotion experiment

	All Workers	Women	Minority Men	White Men
Panel A: Free response, first answer is more school				
See avatars	-0.069**	-0.037	-0.058	-0.073
	(0.034)	(0.040)	(0.082)	(0.156)
Control mean	0.340	0.310	0.349	0.478
Panel B: Free response, first answer is higher quiz score				
See avatars	0.000	-0.053	0.163	-0.119
	(0.040)	(0.048)	(0.102)	(0.165)
Control mean	0.526	0.562	0.482	0.413
Panel C: Free response, first answer is it was random				
See avatars	-0.014	0.005	-0.062*	0.012
	(0.019)	(0.024)	(0.034)	(0.102)
Control mean	0.059	0.058	0.048	0.087
N	691	464	151	76

Note: This table shows the effects of learning that mostly white men were previously promoted 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 other workers evaluated by the algorithm. The outcomes are the same as in Appendix Table A10. 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 levels indicated by *, **, and ***, respectively.
[Click to return to Section 6 (Effects of Perceived Algorithmic Discrimination)]

B Manager recruitment and manager task

Managers were recruited from Prolific. They were required to be white men and employed outside of Prolific. Having all managers be white men serves three purposes: (i) to best proxy for cases in which a women or racial minority man might be the most likely to feel discriminated against, (ii) to minimize noise in the experiment caused by having large variation in manager characteristics, since only 54 managers participated in the promotion experiment and only 20 in the hiring experiment, and (iii) to preserve the sample of possible women and racial minority men to participate in the experiment itself. As illustrated in Figure A3, initially 75 managers were recruited to evaluate the 1800 workers in the historical sample, at which point they were randomly assigned to be a demographic-blind or non-blind manager. They were independently randomly assigned to evaluate workers who scored in the bottom two quintiles of average baseline quiz scores, the second and third quintiles, third and fourth quintiles, or fourth and fifth quintiles. Demographic-blind managers saw workers' education level (no college, college degree, or more than college) and 1-5 stars indicating the workers' approximate quintile of average baseline quiz scores, which are predictive of the workers' proofreading skill. Non-blind managers also saw workers' self-identified race/ethnicity and gender as well as an avatar that workers built to look like themselves in the baseline survey.

Managers were shown examples of the paragraphs that workers would be asked to proofread in the harder job and then chose three out of 24 historical workers to promote, viewing worker profiles that had either just quiz scores and education, or also avatars, race, and gender, depending on the manager's random assignment. The promoted workers' performance in the harder job determined the managers' bonus payment, which represented the majority of their total payment, and thus managers were incentivized to choose the workers they thought would do the best at the harder job. Specifically, managers were paid $1 to complete the survey, which took on average 7 minutes, and on average earned $4 in bonuses based on the performance of the workers in the historical sample that they had promoted.

The same managers were brought back to evaluate the workers in the experimental samples. They were randomly assigned to a group of 120 workers in the promotion (hiring) experiment who fell into the managers' (previously randomly assigned) quiz-score quintile group but had otherwise been randomly grouped together, with one demographic-blind and one non-blind manager assigned to evaluate the workers in each group. Managers saw workers in the same quiz-score quintile groups in both the historical and main samples and the same information about workers in the historical and main samples.

After a reminder about the harder proofreading job and that they had done this worker-selection task before, each manager evaluated workers in groups of forty. In the promotion experiment, they evaluated three sets of forty workers independently, promoting three in each group, and in the hiring experiment, they evaluated nine sets of forty workers independently, hiring one in each group. Managers knew that their decisions for **one** of the three or nine groups would be randomly chosen to be implemented, and the workers they chose in that group would determine their bonus payment. This ensured that every worker was evaluated by both a demographic-blind and non-blind manager, though only the decisions of one—or (one of) the algorithm(s)—were implemented. This generated counterfactual data on how workers would have been assigned to jobs under all three (four) procedures, which were used in the analysis (but not in the implementation of the experiment). In the promotion experiment, managers were paid $2 to complete the survey, which took on average 10 minutes, and on average earned $4 in bonuses based on their chosen workers' performance. In the hiring experiment, managers were paid $2 to complete the survey, which took on average 15 minutes, and on average earned $3.25 in bonuses based on their chosen workers' performance.

Of course, there was attrition between the original manager recruitment and their return to evaluate the main sample workers. Initially 75 managers were recruited in order to account for up to 20 percent attrition. Then, after the promotion experiment baseline survey was complete, one demographic-blind and one non-blind manager were each matched with a random group of 120 workers in their same (randomly assigned

to managers) quiz-score quintile group. After several days and several reminders, when a manager failed to return to evaluate the new workers, they were replaced with a randomly chosen manager who matched their random assignment (demographic-blind or not $\times$ quiz-score quintile group) from the group of managers that had not been assigned to evaluate workers initially. This continued until all workers had been evaluated. 54 managers participated in the promotion experiment.

The same managers were invited back for the hiring experiment, now eight months following their initial recruitment. Again, one demographic-blind manager and one non-blind manager was each matched with a random with a group of 360 workers in their same (randomly assigned to managers) quiz-score quintile group. The number of groups of workers per manager was larger so that fewer managers would be needed, assuming there would be major attrition given the amount of time since the initial survey. This time, managers were matched with groups randomly but those who had returned to evaluate workers in the promotion experiment (a signal that they would return again) and had previously promoted mostly white men (in order to try to minimize variation in that dimension of treatment) had a higher chance of being initially paired with workers and offered the worker-evaluation task. The same replacement procedure was followed when managers did not return after several days and reminders. 20 managers participated.

In one group of 120 workers in the promotion experiment, I ran out of managers who could be paired with the group because none of the eligible managers returned. I invited two managers who had already evaluated a group of 120 workers in the same quiz-score quintile group to return and evaluate another set. However, only one returned. In the hiring experiment, in one group of 360, one of the initially-assigned managers returned but none of the other eligible managers returned. In both experiments, I implemented the decisions of the one manager who returned and paid them accordingly, but the workers are dropped from the experimental sample. Results are fully robust to their inclusion in the promotion experiment (in the hiring experiment, only the worker who was hired, not the ones who were not, were offered the corresponding survey due to budget constraints and the fact that they could not be included in the main sample). These results are available upon request.

C Training the algorithm

The demographic-blind algorithm is a regression-based random forest model that uses workers' average quiz-score quintile fixed effects and education fixed effects (no college, college degree, more than college) to predict which workers will do the best if they are promoted. The algorithm used these variables to estimate a model predicting good performance in the harder job. As described in Section 2.2.1, the harder job involved proofreading 12 paragraphs from articles published in leading scientific journals.

The algorithm predicted the average number of mistakes correctly highlighted minus the number of non-mistakes incorrectly highlighted, each standardized to have mean zero and standard deviation one. Since workers could choose to stop proofreading at any time in the harder job, using the total performance across all 12 paragraphs also includes a component of retention. For the purposes of the experiment, the accuracy and predictiveness of the algorithm were irrelevant – what mattered was its inputs and that it predicted performance, as this is what was communicated to workers.

The non-blind algorithm, used only in the hiring experiment, simply adds interactions of each of the above predictors with race and gender fixed effects. Both algorithms are unbiased using calibration: conditional on predicted performance, race and gender are not independently predictive of performance in the training sample or the small number of promoted/hired workers for whom I observe both predicted performance and performance. If anything, the non-blind algorithm promotes *more* racial minorities in the experimental sample.

The algorithms were trained on a sample of 500 workers recruited from MTurk using Cloud Research's pre-approved participant pool. The sample for the entire study was originally was going to be recruited from MTurk, since these types of tasks are more common there. However, in recruiting this training sample, it became evident that it would be impossible, or at least prohibitively expensive, to recruit enough racial minority participants from CloudResearch's pre-approved pool of MTurk workers. Like the experimental sample, women and racial minorities are also over-represented in the training sample: 33 percent are white women, 18 percent are non-white women, 18 percent are non-white men, and 30 percent are white men. There are no significant race or gender differences in performance in the training sample. These workers completed (in one survey) both the baseline survey *and* paragraphs the proofreading job.

The random forest model was used to predict worker performance in the harder job after completing only the baseline survey in the historical and two experimental samples. Like the managers, the algorithm selected three workers in each group of 24 (in the historical sample) or 40 (in the promotion experiment sample) to do the harder job. In the hiring experiment, again like the managers, the algorithm selected one worker in each group of 40. In each group, the algorithm ranked workers by their predicted performance, and assigned the top three or one worker(s) to the harder job. Ties were resolved randomly (e.g. if four workers tied for the second-highest predicted performance, two were randomly assigned to the higher-status job along with the worker with the highest predicted performance, etc.). Similarly to how the managers actually evaluated all worker groups in order to generate counterfactual data to use in the analysis, the algorithm's "decisions" were determined for all workers, not just those in groups randomly assigned to have the algorithms' decisions implemented.

D Variable definitions

Unless otherwise specified, definitions apply in both the promotion experiment and hiring experiment. The measures of perceived discrimination are described in full detail in the text and so are not included here.

D.1 Outcome variables

Retention. (Promotion experiment only). After the sixth paragraph, workers had the option to skip to the end of the survey after proofreading each paragraph. The measures of retention are the number they proofread, whether they do more than six (i.e. don't quit right away), and whether they do all eighteen. The question about skipping to the end of the survey was the following: "Would you like to continue to the next paragraph? You've done the required six paragraphs to receive your participation payment. Remember, you can earn $0.25 per paragraph that you do a good job proofreading. There are X more paragraphs to proofread. There is no penalty for stopping now, but you can only be paid for paragraphs that you complete."

- "I'd like to proofread another paragraph"
- "I do not want to proofread any more paragraphs. (By selecting this option, you will continue to the end of the survey)"

Effort and performance. (Promotion experiment only). The effort measures come from data collected passively while workers do the proofreading job. This includes the number of times they click on the page per paragraph, the amount of time they spend on each paragraph, how many words they correctly highlight that are mistakes, how many words they incorrectly highlight that are not mistakes, and the bonus they earn.

Future labor supply. All reservation wages are elicited using a multiple price list. Workers are told that in each case, one of the wages/wage schedules in the rows below would be randomly selected and their answer in that row would determine whether they would be evaluated in the future and offered the corresponding wage/job depending on their evaluation, and that workers who said they were interested at a given wage may be randomly selected to be offered the job if interest exceeds the number of workers we needed to hire. *Note that there are many long blocks of text below; workers were always shown text in shorter blocks and with key words bolded for emphasis which has been removed here.*

Promotion experiment

- *Same procedure:* "We will offer some workers a chance to be evaluated again and assigned to more proofreading tasks like this in the future. In one future round, the mechanism used to assign people to the easier, lower-paying task versus the harder, higher-paying task will be the same as today:"

  In the demographic-blind manager arm: "a manager will review your average quiz score and education and assign you to proofread science articles for kids or articles published in leading scientific journals"

  In the non-blind manager arm: "a manager will review your applicant profile (avatar, average quiz score, etc.) and assign you to proofread science articles for kids or articles published in leading scientific journals"

  In the algorithm arm: "an algorithm will use your average quiz score and education to determine which task you would be assigned to"

[All arms] "You can earn twice as much per hard paragraph than easy paragraph. At what wages would you be interested in being evaluated again under this task assignment mechanism?"

- *Cutoff rule:* "In another future round, we will use screening quiz scores to determine who is assigned to the easier, lower-paying task vs the harder, higher-paying task: Among workers interested in the job, the top scorers on the screening quizzes will be offered the harder job. You can earn twice as much per hard paragraph than easy paragraph. At what wages would you be interested in being evaluated again under this task assignment mechanism?"

  Multiple price list in both cases:

	I would **not** want this job (at these wages)	I would want this job (at these wages)
Earn **$0.05** for each high-quality easy paragraph; **$0.10** for each high-quality hard paragraph	☐	☐
Earn **$0.10** for each high-quality easy paragraph; **$0.20** for each high-quality hard paragraph	☐	☐
...	**...**	**...**
Earn **$0.50** for each high-quality easy paragraph; **$1.00** for each high-quality hard paragraph	☐	☐

  The calculation of the outcome variables from the multiple price lists are described in detail in the paper; the overall reservation wage is a function of the point at which they switch from not wanting the job to wanting the job and the decomposed reservation wage imagines that each worker saw an individualized multiple price list with their expected wages, given their beliefs about the probability that they would be assigned to the harder, higher-paying task.

Hiring experiment:

- *Same procedure:* "We will offer some workers a chance to be evaluated again and potentially hired for the proofreading task in the future. In one future round, workers will be evaluated and hired the same way as today:"

  In the demographic-blind manager arms, regardless of whether they see previous hires' avatars, race and gender: "a manager will review your and others' average quiz scores and education again and decide who to hire for the proofreading task. The outcome may be different from today (you might be hired this time) because they would be reviewing a different set of workers."

  In the non-blind manager arm: "a manager will review your and others' applicant profiles (avatar, average quiz score, etc.) and decide who to hire for the proofreading task. The outcome may be different from today (you might be hired this time) because they would be reviewing a different set of workers."

  At what wages would you be interested in being evaluated again under this task assignment mechanism?"

- *Cutoff rule:* "In another future round, we will use screening quiz scores to determine who is hired to do the proofreading task: Among workers interested in the job, the top scorers on the screening quizzes will be offered the harder job. At what wages would you be interested in being evaluated again under this task assignment mechanism?"

  Multiple price list in both cases:

	I would **not** want this job (at these wages)	I would want this job (at these wages)
Earn **$0.10** for each high-quality paragraph	☐	☐
Earn **$0.20** for each high-quality paragraph	☐	☐
...	...	...
Earn **$1.00** for each high-quality paragraph	☐	☐

  The reservation wage is the midpoint between the two wages at which the worker switches from not wanting the job to wanting the job.

Cooperation and sharing with managers. (Promotion experiment only, manager sample only). This section of the survey started with, "The next three questions are going to ask you about whether you would want to work together with your manager on a task in the future and how you would want to share a thank-you bonus with your manager. So, we'll give you a reminder about what you know about your manager." Workers then saw the same manager profile and three previously-promoted workers from the beginning of the survey (Appendix Figure A6).

- **Cooperative task reservation wage.** Workers were told the following: "In other future jobs, we will also be asking workers to work together with their manager from today's task or a similar manager to produce high-quality summaries of the more complicated scientific texts. Instead of using managers to assign workers to the harder summarizing task: (1) All workers will summarize complicated scientific paragraphs (summarizing round) (2) Managers will review worker summaries, leave comments and choose a bonus payment for the worker (3) Workers will have another chance to revise their work (editing round). The manager will have some discretion over how much workers are paid, but there will be a base payment per paragraph. Below, we would like to know if you would be interested in the job at each of the given base payments per paragraph, where this base payment applies to both to first summarizing round and the second editing round. We will randomly choose one of the base payments below, and among the workers who said they would be interested in the job at that wage, we will randomly choose 20 workers to be offered the job." The reservation wage is the midpoint between the two wages at which the worker switches from not wanting the job to wanting the job.

	I would **not** want this job (with this base payment)	I would want this job (with this base payment)
Earn **$0.05** for each high-quality summary and edit	☐	☐
Earn **$0.10** for each high-quality summary and edit	☐	☐
...	...	...
Earn **$1.00** for each high-quality summary and edit	☐	☐

- **Willingness to pay to choose own manager.** Next, workers are told the following: "Now, imagine that we will pay $1.00 per high-quality completed summary and you are interested in doing this future task at this wage. The default in this future task is that you will be assigned to the same or a similar manager as in the survey today. This manager will review your summaries, leave comments, and suggest your bonus payment. But, you can give up part of your wage in order to be able to choose who you want to review your work from a list of 5 managers, including the one who assigned you today. Below, we would like to know if you would want to keep your same manager or pay to choose your manager, if it costs the amount of money in the leftmost column to choose your own manager. We will be randomly choosing 20 additional workers who are interested in this job if we pay $1.00 per high-quality summary in the first round and edit in the second round, and offer them this job. We will randomly choose one of the prices below and implement their choices at that price. For example, if we choose the price $0.10, and you said that you would want to choose your own manager at that price, you would get to choose your own manager, and be paid $0.90 per high-quality summary. Otherwise, you will work with the same manager as today and be paid $1.00 per high-quality summary. In either case, your manager would determine your bonus payment." The willingness to pay is the midpoint between the two prices at which the worker switches from wanting to switch to not wanting to switch.

	I would **keep** the manager from today (at this price)	I would want to choose a new manager (at this price)
Pay **$0.05** per summary/edit to choose a manager	☐	☐
Pay **$0.10** per summary/edit to choose a manager	☐	☐
...	...	...
Pay **$0.95** per summary/edit to choose a manager	☐	☐
Pay **$1.00** per summary/edit to choose a manager	☐	☐

- **Generosity.** Finally workers play a standard dictator game with their manager. They are told, "Finally, as a thank-you for your participation in the study, 20 workers will be randomly selected to receive an extra $20 bonus. If you are chosen, you have the option of allocating some of your thank-you bonus to your manager, whose job was to assign you and other workers to different tasks. If you are selected as to receive the extra $20 bonus, how much, if any, of your extra $20 bonus would you share with your manager? If you are selected to receive the extra bonus, your choice here will be implemented. (Your manager will not know who did or did not share thank-you bonuses with them.)"

  Workers chose how much they would like to keep on a sliding scale, with labels "You keep nothing, Manager gets $20" at 0, "You keep $10, Manager gets $10" at 10, and "You keep $20, Manager gets nothing" at 20. The outcome is the chosen amount they share.

Beliefs about promotion. After the reservation wage elicitations for the case where they would be evaluated by the same procedure as in the experiment, they were asked, "if you are selected to participate in this round, what do you think is the probability that [your manager/the algorithm] ... will assign you to the higher-paying, harder task? (in the promotion experiment) ...would choose you to do the proofreading task? (in the hiring experiment).

After the reservation wage elicitation for the case where they would be evaluated using the cutoff rule in baseline quiz scores, the question was instead "If you are selected to participate in this round, what do you think is the probability that you would be among the highest scorers on the quizzes and therefore ... assigned to the higher-paying, harder task? (in the promotion experiment) ... hired to do the proofreading task (in the hiring experiment).

In both cases, workers answered by dragging a sliding scale from 0 to 100, labeled as "I will definitely be assigned to the lower-paying, easier task" at 0, "It's the same as a coin flip" at 50, and "I will definitely be assigned to the higher-paying, harder task" at 100. The outcome is workers' reported probability.

Comprehension of evaluation procedures. At the end of the hiring experiment, workers learned that 100 workers would be randomly selected to earn a small bonus for each of two comprehension questions they answered correctly.

- **Decision-maker inputs.** In both the manager and algorithm samples, workers were asked to select all that apply to the question "What information did [your manager/the algorithm] have about workers when deciding who to hire for the proofreading job? The list below was in a random order.

  Age; *Gender;* *Race/Ethnicity;* *An avatar;* All three numeric quiz scores from the baseline survey (science, spelling, and grammar); The average quiz score (for example, 85%); *1-5 stars representing average quiz scores;* Work history; Time on Prolific; *Education*

  Workers were correct if they selected only the italicized and italicized+underlined options above in the non-blind decision-maker arms, and if they selected only the italicized (without an underline) options in the demographic-blind decision-maker arms.

- **Manager incentives (payment structure).** In the manager sample, workers were asked to select all that apply to the question, "What as the basis of your manager's bonus payment?" The list below was in a random order.

  The performance in the proofreading task of the workers they hired; How many workers they hired; Whether they choose workers with the highest screening quiz scores; Whether they chose diverse workers; *How many workers they hired showed up to do the proofreading task*

Workers were correct if they only selected the italicized options in any manager treatment arm

- **Algorithm design.** In the algorithm sample, workers were also to select all that apply to the question, "What do you know about how the algorithm was designed?"

  It was predicting worker performance at the proofreading task; It was predicting whether a manager would have hired a worker; *Call the thing the algorithm was predicting Y. The algorithm hired whoever had the highest predicted Y; Call the thing the algorithm was predicting Y. The algorithm would provide the same predicted Y for any worker with the same education and quiz score;* Call the thing the algorithm was predicting Y. The algorithm could favor certain people and hire them even if they didn't have the highest predicted Y.

  Workers were correct if the selected only the italicized and italicized+underlined options above in the demographic-blind decision-maker arms, and if they selected only the italicized (without an underline) options in the non-blind decision-maker arms.

Psychological mechanisms. Note: before these questions were asked, workers were assured that their answers to them were confidential and would not affect their evaluation or chances of future work.

- **Affect.** Workers are asked to indicate to what extent they feel each of the following emotions right now, on a scale from 1 (not at all) to 6 (very much): happy, at ease, anxious, annoyed, motivated, calm, tired, bored, gloomy, active. This is the standalone short- form 10-item Daniels five-factor measure of affective well-being (D-FAW; Russell and Daniels, 2018). Mixed in with the standard items, they are also asked how discouraged and upset they are in order to validate responses to how motivated and at ease they are, key emotions of interest. I standardize their rating of each emotion to have mean zero and standard deviation one in the demographic-blind manager group and the overall index of psychological well-being is the mean of the twelve standardized variables, with the sign of negative emotions (anxious, annoyed, tired, bored, gloomy, discouraged, upset) flipped.

- **Self-efficacy.** Work self-efficacy is one's confidence in their ability to do the tasks required of them in a particular job. To assess participants' work self-efficacy about the tasks at hand, workers are asked how much they agree or disagree with the following statements on a Likert scale from 1 (Strongly disagree) to 5 (Strongly agree):

  (1) I am capable of doing the harder proofreading job well; (2) I would have liked a chance to do the harder proofreading task; (3) I am confident in my ability to work under pressure; (4) I am capable of doing the easier proofreading job well; (5) I did a good job on the proofreading task today; (6) I was able to improve as I proofread more paragraphs

  And to understand their self-efficacy related to the the underlying skills they posses, they are asked to indicate their skill level in the following areas on a Likert scale from 1 (Not at all skilled) to 5 (Completely skilled):

  (1) Written communication; (2) Oral communication; (3) Problem solving; (4) Numeracy; (5) Motivation; (6) Learning new material

  For both measures, I standardize workers' responses to each component to have mean zero and standard deviation one in the control group and take the average to form an index of task-specific and skills-based self-efficacy. In the promotion experiment, I average both together as a total index in Appendix Table A11, but the (lack of) effects on the two indices separately are similar. In the hiring experiment, I only observe the index of underlying skills.

- **Self-reported interest in future work.** Workers were asked to indicate how much they agreed or disagreed with the following on a Likert scale from 1 (Strongly disagree) to 5 (Strongly agree):

  (1) I want to complete more surveys for this employer; (2) I want to complete more surveys with tasks that are assigned in this way; (3) I want to complete more proofreading tasks of this level of difficulty; (4) I want to complete more difficult proofreading tasks.

  The outcomes in columns 1-4 of Appendix Table A8 are indicators for whether workers strongly agreed with each of the above statements. The outcome in column 5 comes from a question in which they were briefly reminded of their evaluation and non-promotion and asked, "Would you have preferred to be assigned to the harder, higher-paying task?" The outcome of interest is an indicator for answering "Definitely" or "I think so," rather than "I don't think so," "Definitely not," or "I don't care." These questions were only asked in the promotion experiment.

- **Job satisfaction.** Workers were asked to indicate how satisfied they were with the following on a Likert scale from 1 (Very dissatisfied) to 4 (Very satisfied):

  (1) How satisfied are you on the whole with the work that you were offered in this survey?; (2) How satisfied are you on the whole with the work that is available on Prolific?; (3) How satisfied are you on the whole with the work that you do outside of Prolific (if applicable)?

  The measure of job satisfaction is an indicator for whether they said that they were very satisfied with the work they were offered in this survey. These questions were only asked in the promotion experiment.

D.2 Control variables

All control variables are the same in the promotion and hiring experiments.

Demographic controls.

- Education: *What is the highest level of education you have attained?* I control for seven mutually-exclusive groups {Less than high school, high school graduate, some college but no degree, 2 year college degree, 4 year college degree, Professional or Masters degree, Doctorate}

- Income: *What is your annual household income (pre-tax)?* I control for nine mutually-exclusive groups {Less than $20,000, $20,000-40,000, $40,000-60,000, $60,000-80,000, $80,000-$100,000, $100,000-120,000, $120,000-140,000, $140,000-160,000, More than $160,000}

- Married: *Are you married or in a long-term partnership?* I control for an indicator for answering "Yes" rather than "No"

- Kids: *Do you have children?* I control for an indicator for answering "Yes" rather than "No"

- Age: *How old are you?* I control for six mutually-exclusive groups {18-24, 25-34, 35-44, 45-54, 55-64, over 65}

- Race: *With what racial group do you most strongly identify?* **and** *Do you identify as Hispanic or Latino?* I code workers as one of Asian, Black, white and Hispanic, or white and non-Hispanic and control for these four mutually-exclusive categories. Workers who identified as multiracial, other

race, or Native American were grouped with Asian, Black, or Hispanic workers based on the race recorded in Prolific's administrative data, which matched one of those groups. Black, Hispanic workers were grouped with Black, Non-Hispanic workers.

- Gender: *What is your gender?* I control for an indicator for answering "Man" rather than "Woman," "Non-binary," or "Other"

Baseline quiz scores. I control for the number of questions workers got correct on each of the following three quizzes:

- Spelling quiz: Workers had 12 seconds to play an audio recording of the word and type it into a text box. They earned a bonus of $0.025 for each word they spelled correctly. The words were: hypothesis, paleontology, equilibrium, herbivore, aerobic, enzyme, homeostasis, chlorophyll, sedimentary, and vertebrae. They were reminded to have their audio on and volume up before beginning, and had two practice words to make sure the audio was working.

- Grammar quiz: Workers had 4 minutes to answer as many of 15 questions as they could. They earned a bonus of $0.04 for each question they answered correctly.

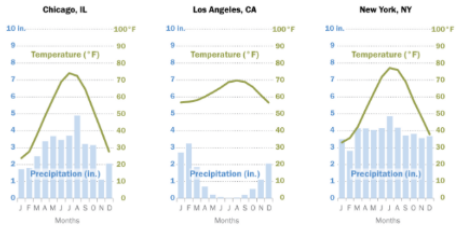
  - Which of these is not a word or phrase? [a lot, *alot*, allot]
  - You've probably heard the phrase "i before e except after c" ...but which of these words defies this rule? [species, science, policies, *all of them*]
  - A plural subject needs [a singular verb, *a plural verb*]
  - When two singular subjects are connected by 'or,' use [*a singular verb*, a plural verb]
  - Which phrase is incorrect? [should've, should have, *should of*]
  - If an opinion-adjective and a fact-adjective are used before a noun, which comes first? [a fact-adjective, *an opinion-adjective,* it doesn't matter]
  - Fill in the blank: "Bad weather can _____ people's ability to work" [*affect*, effect]
  - Every sentence must have a subject and [an object, *a verb,* an adjective, an adverb, a phrase]
  - If someone says "I'm sorry" you can ____ their apology [except, *accept*]
  - The order of a basic positive sentence is [*subject-verb-object*, verb-object-subject, object-verb-subject, subject-object-verb]
  - Which is correct? [*they're looking good*, their looking good, there looking good]
  - Fill in the blank. "____ so hot outside!" [*it's*, its]
  - Which is correct? [the boy's dog bark loudly, the boy's dog loudly barks, *the boy's dog barks loudly*, they boys dog barks loudly]
  - Select all of the sentences that are written in the active voice [grass is eaten by cows, the books were written by that author I like, *she drove over the bridge, horses eat hay, they clued us in*]
  - Select all that apply: The contraction "she's" can mean... [*she is, she has,* she was]

- Science quiz: Workers had 4 minutes to answer as many of 11 questions as they could. They earned a bonus of $0.05 for each question they answered correctly.

When large areas of forest are removed so land can be converted for other uses, such as farming, which of the following occurs?

- Decreased carbon dioxide
- Colder temperature
- **Increased erosion**
- Greater oxygen production

These graphs show the monthly precipitation and average temperature for three cities in the United States over the course of one year. Based on the graphs, which city has the greatest annual range of temperatures?



- **Chicago**
- Los Angeles
- New York
- The all have the same annual temperature range

Which of these is a major concern about the overuse of antibiotics?

- **It can lead to antibiotic-resistant bacteria**
- There will be an antibiotic shortage
- Antibiotics can cause secondary infections
- Antibiotics will get into the water system

Which of the following is an example of genetic engineering?

- Growing a whole plant from a single cell
- Finding the sequences of bases in plant DNA
- **Attaching the root of one type of plant to the stem of another type of plant**
- Inserting a gene into plants that makes the plant resistant to insects

Many diseases have an incubation period. Which of the following best describes what an incubation period is?

- The recovery period after being sick
- **The period during which someone has an infection, but is not showing symptoms**
- The period during which someone builds up immunity to a disease
- The effect of a disease on babies

Oil, natural gas, and coal are examples of:

- Geothermal resources
- Biofuels
- **Fossil fuels**
- Renewable resources

An antacid relieves an overly acidic stomach because the main components of antacids are ...

- Acids
- Neutral
- Isotopes
- **Bases**

A scientist is conducting a study to determine how well a new medication treats ear infections. The scientist tells the participants to put 10 drops in their infected ear each day. After two weeks, all participants' ear infections had healed.

Which of the following changes to the design of this study would most improve the ability to test if the new medication effectively treats ear infections?

- Create a second group of participants with ear infections who use 15 drops a day
- Have participants put ear drops in both their infected ear and healthy ear
- Have participants use ear drops for only one week
- **Create a second group of participants with ear infections who do not use any ear drops**

The time a computer takes to start has increased dramatically. One possible explanation for this is that the computer is running out of memory. This explanation is a scientific...

- Experiment
- **Hypothesis**
- Observation
- Conclusion

What is the main cause of seasons on the Earth?

- **The tilt of the Earth's axis in relation to the sun**
- The distance between the Earth and the sun
- Changes in the amount of energy coming from the sun
- The speed that the Earth rotates around the sun

A car travels at a constant speed of 40 miles per hour. How far does the car travel in 45 minutes?

- 25 miles
- **30 miles**
- 35 miles
- 40 miles

Previously-selected worker education and baseline performance. In all treatment groups, workers saw the education and baseline performance (as 1-5 stars representing quintiles of average quiz scores) of three workers who were previously promoted or hired by their manager or the algorithm that evaluated them. Recall that workers were in "quiz-score groups" that only included workers within one quiz-score quintile of each other, and I include controls for the quiz-score group in all regressions. I also include controls that measure how a worker compares to the previously-promoted workers, conditional on their quiz-score group:

- Performance: I control for four mutually-exclusive indicators: Whether they see three previously-selected workers with the maximum number of stars, two previously-selected workers with the maximum number of stars (and thus one with the minimum), one worker with the maximum number of stars (and thus two with the minimum), or three previously-selected workers with the minimum number of stars.
- Education: I control for whether they see three previously-selected workers with more education than themselves, two with more education than themselves, one with more education than themselves, three with the same education as themselves, one with less education than themselves, two with less education as themselves, or three with less education than themselves

D.3 Variables used in heterogeneity

Past experiences of discrimination. Workers answered a module on their prior work history at the end of the baseline survey. They are told that the remaining questions are about if they've ever felt treated unfairly at work because of their age, race, gender, etc., and asked:

- Do you think you have ever been discriminated against while applying for a job? That is, do you think you were or weren't hired unfairly because of some characteristic like your race, age, gender, disability status, religion, etc. rather than because of your qualifications?
- Do you think you have ever been discriminated against by your employer in a promotion decision? That is, do you think you were or weren't promoted unfairly because of some characteristic like your race, age, gender, disability status, religion, etc. rather than because of your qualifications?
- Do you think you have ever been discriminated against by your employer in a termination decision? That is, do you think you were or weren't fired unfairly because of some characteristic like your race, age, gender, disability status, religion, etc. rather than because of your qualifications?
- During day-to-day activities at work or while working, do you feel that you have been treated differently than others based only on your gender, race, age, or some other demographic characteristic?

After each question, they could answer "Yes," "No," or "It's impossible to know." If they answered yes, they were asked if they thought they were discriminated against based on gender, race, age or something else, and to select all that apply, including an option to write in "other." Finally, they were asked if they experienced positive or negative discrimination: e.g. "**I was fired** because of some demographic characteristic and **I shouldn't have been**," or "**I was not fired** because of some demographic characteristic and **I should have been fired.**"

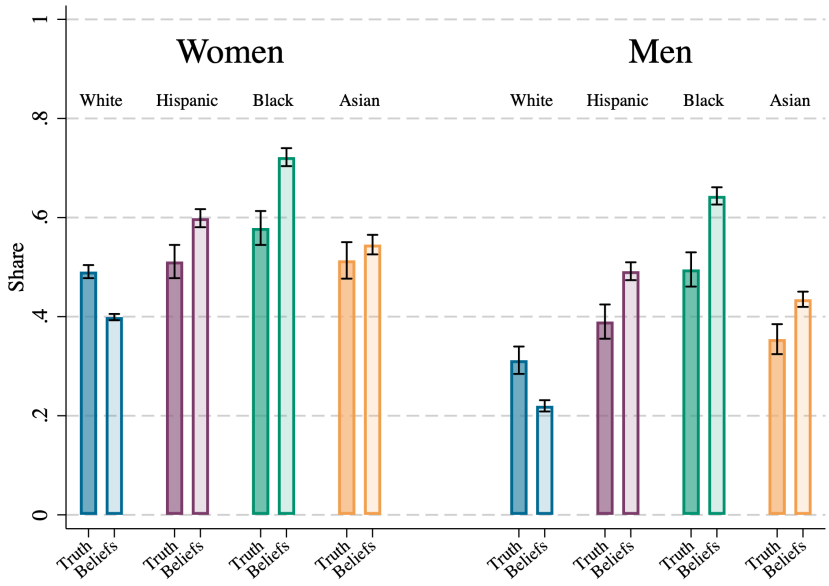
In the heterogeneity analysis, I code someone as having experienced discrimination in the past if they say yes to any of the four questions and say that the discrimination was on the basis of race and gender and say

that it was negative discrimination (almost none of the workers report experiencing positive discrimination, likely because they did not think of that as discrimination in the first place and so were not asked whether it was positive or negative).

Confidence. After each of the quizzes, workers were asked how many questions they thought they got right from a multiple choice list from 0-10, 11, or 15 in the case of spelling, science, and grammar, respectively. They earned a $0.20 bonus if they were correct about how many questions they got right for each quiz. The difference from their actual score is a measure of domain-specific confidence. As a general measure, I standardize the difference between their belief and their score to have mean zero and standard deviation one in the demographic-blind manager group, average the standardized variables across the three quizzes, and use an indicator for having this statistic be above-median to test whether confidence matters in heterogeneity analysis.

Beliefs about the prevalence of discrimination. Directly after the questions about past experiences of discrimination, I ask workers, "What fraction of each of the following demographic groups they think would answer "yes" to any of the previous questions? That is, what percent of each demographic group do you think reports that they have been discriminated against while searching for a job, in promotion or termination decisions, or in day-to-day work?" I tell them that the sample is other workers recruited on Prolific. They answer on a slider for the eight {Asian, Black, Hispanic, White} $\times$ {men, women} groups from 0 to 100, labeled at 0 and 100 "No one in this group reports experiencing discrimination" and "Everyone in this group reports experiencing discrimination," respectively.

Their answer is incentivized; I randomly select one group and workers are paid a $0.50 bonus if their answer is within 5pp of the truth. The measure used in heterogeneity analysis is an indicator for whether their beliefs about the prevalence of discrimination for their *own group* is above- or below-median. The figure below shows the average belief compared to the truth separately for workers in each group.



Note: This figure plots the fraction of workers in each race $\times$ gender group who report ever experiencing discrimination at work in the past and the average belief about workers in that group about the same statistic. Standard errors are shown in black bars.

Racial- and gender-identity centrality. Workers completed two standard scales that measure racial- and gender-identity centrality, or the degree to which race and gender are, respectively, important to one's sense of self. In both cases, workers were asked to rate statements on a Likert scale from 1 (Strongly disagree) to 7 (Strongly agree). Indices are formed by standardizing each response to have mean zero and standard deviation one and taking the average, flipping the items marked with a (-) to have the opposite sign.

- Racial identity centrality:
  - Overall, my racial group identity has very little to do with how I feel about myself (-)
  - In general, identifying with my racial group is an important part of my self-image (+)
  - My destiny is tied to the destiny of other people in my racial group (+)
  - The racial group I identify with is unimportant to my sense of what kind of person I am (-)
  - I have a strong sense of belonging to my racial group (+)
  - I have a strong attachment to other people in my racial group (+)
  - The racial group I identify with is an important reflection of who I am (+)
  - Belonging to the racial group I identify with is not a major factor in my social relationships (-)

- Gender identity centrality:
  - I often think about the fact that I am a [man/woman] (+)
  - Overall, being a [man/woman] has very little to do with how I feel about myself (-)
  - The fact that I am a [man/woman] rarely enters my mind (-)
  - I am not usually conscious of the fact that I am a [man/woman] (-)
  - In general, being a [man/woman] is an important part of my self-image (+)
  - Being a [man/woman] is an important reflection of who I am (+)
  - In my everyday life, I often think about what it means to be a [man/woman] (+)

E Instrumental variables (IV) specification results

As described in Section 3.1, an IV strategy estimating the effects of perceived discrimination has interpretation challenges due to the amorphous nature of perceived discrimination. The general exclusion restriction—that treatment only affects outcomes via "perceived discrimination"—is valid, because the experimental design carefully eliminates any other channels. That said, IV estimates effectively scale the ITT estimates by the "first stage," in this case, the effect of treatment assignment on perceived discrimination.

I measure many (highly correlated) measures of perceived discrimination, but they vary substantially in the size of the first stage. Thus, IV estimates vary greatly depending on which measure of perceived discrimination is used. For example, the first stage for the main measure of perceived discrimination is 31pp, but the first stage for the measure based on whether workers make a complaint about bias is only 10pp. One can think of the frequency of a variable as the inverse of its seriousness or severity, so it is expected that these two measures are different in this way.

Interpreting an IV estimate in this case requires taking a stand on the "type" of perceived discrimination through which the treatment has effects, and assuming away treatment heterogeneity depending on the severity of the perceived discrimination. For example, if only those who make a complaint about bias are affected, the implied effect of perceived discrimination is 3.3 times as large as the case where anyone who thinks demographics played a role in their evaluation is affected.

Thus, the paper presents the ITT estimates. These are a lower bound on the average effects of perceived discrimination in the case that everyone perceives some discrimination. Here, I present the results from an IV specification that instruments the main measure of perceived discrimination with treatment assignment. The main measure of perceived discrimination has the largest first stage of all of the measures of perceived discrimination, making these the most conservative IV estimates using any observed measure.

Table E1: The effect of perceived discrimination (IV estimates) on retention

	Num paragraphs	More than 6	Did all 18
	(1)	(2)	(3)
Panel A: Manager sample, promotion experiment			
Perceived manager bias	-1.596**	-0.050	-0.156*
	(0.806)	(0.056)	(0.081)
N	1387	1387	1387
Control mean	15.916	0.915	0.776
Panel B: Algorithm sample, promotion experiment			
Perceived algorithmic bias	0.820	0.169	0.029
	(2.266)	(0.170)	(0.226)
N	691	691	691
Control mean	15.456	0.881	0.730

Note: This table shows instrumental variables estimates of the effect of the main measure of perceived discrimination on retention, instrumented with random assignment to treatment. This effectively transforms the ITT effects in Table 2 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Figure 1). Outcomes and control variables are otherwise the same as Table 2. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table E2: The effect of perceived discrimination (IV estimates) on effort and performance

	Minutes (1st 6)	Clicks (1st 6)	Correct (1st 6)	Incorrect (1st 6)	Bonus (1st 6)
	(1)	(2)	(3)	(4)	(5)
Panel A: Manager sample, promotion experiment					
Perceived manager bias	-0.085	1.975	1.003	0.698	0.023
	(0.175)	(5.652)	(0.719)	(0.687)	(0.068)
N	1389	1389	1389	1389	1389
Control mean	4.957	36.804	14.817	3.422	0.919
Panel B: Algorithm sample, promotion experiment					
Perceived algorithmic bias	0.356	-10.615	-4.137*	1.321	-0.348*
	(0.481)	(15.439)	(2.238)	(1.612)	(0.206)
N	691	691	691	691	691
Control mean	4.817	37.941	15.016	3.679	0.935

Note: This table shows instrumental variables estimates of the effect of the main measure of perceived discrimination on effort and performance, instrumented with random assignment to treatment. This effectively transforms the ITT effects in Table 3 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Figure 1). Outcomes and control variables are otherwise the same as 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 E3: The effect of perceived discrimination (IV estimates) on future labor supply

	Res. wage (same)	WTP cutoff
	(1)	(2)
Panel A: Manager sample, promotion experiment		
Perceived manager bias	0.072**	0.032**
	(0.032)	(0.014)
N	1338	1325
Control mean	0.265	-0.014
Panel B: Algorithm sample, promotion experiment		
Perceived algorithmic bias	0.066	0.007
	(0.086)	(0.033)
N	672	660
Control mean	0.282	-0.008

Note: This table shows instrumental variables estimates of the effect of the main measure of perceived discrimination on reservation wages, instrumented with random assignment to treatment. This effectively transforms the ITT effects in Table 4 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Figure 1). Outcomes and control variables are otherwise the same as Table 4. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table E4: The effect of perceived discrimination (IV estimates) on beliefs about promotion

	Pr(Promoted|Same)	Pr(Promoted|Cutoff)	Diff Pr(Promoted)
	(1)	(2)	(3)
Panel A: Manager sample, promotion experiment			
Perceived manager bias	-8.668**	-3.711	-4.957**
	(3.675)	(3.831)	(2.298)
N	1385	1385	1385
Control mean	47.219	48.491	-1.272
Panel B: Algorithm sample, promotion experiment			
Perceived algorithmic bias	-2.722	-8.743	6.021
	(9.552)	(10.455)	(5.582)
N	689	689	689
Control mean	45.908	48.201	-2.293

Note: This table shows instrumental variables estimates of the effect of the main measure of perceived discrimination on beliefs about promotion, instrumented with random assignment to treatment. This effectively transforms the ITT effects in Table 6 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Figure 1). Outcomes and control variables are otherwise the same as Table 6. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table E5: The effect of perceived discrimination (IV estimates) on sharing with and avoidance of manager

	Cooperative task RW	WTP new manager	Shared zero	Amt bonus shared
	(1)	(2)	(3)	(4)
Perceived manager bias	0.025	0.124**	0.158*	-0.815
	(0.046)	(0.060)	(0.091)	(0.733)
N	1349	1030	1383	1383
Control mean	0.341	0.223	0.522	2.857

Note: This table shows instrumental variables estimates of the effect of the main measure of perceived discrimination on sharing and willingness to work with managers, instrumented with random assignment to treatment. This effectively transforms the ITT effects in Table 5 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Figure 1). Outcomes and control variables are otherwise the same as Table 5. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

F Other heterogeneity

In the main analysis, I focus on heterogeneity by race and gender as it yields the most consistent patterns of the pre-registered heterogeneity analysis. Here, I test for heterogeneity along other dimensions.

Retention and performance. Appendix Tables F1 and F2 show results for these outcomes from a single multivariate regression that simultaneously tests gender and all other pre-registered dimensions of heterogeneity (an indicator for whether an individual reported experiencing discrimination at work in the past, has above-median beliefs about the prevalence of discrimination against their group, or has above-median confidence in related domains). The effects on retention are be most negative for workers with below-median confidence. The pattern of gender heterogeneity persists strongly.

I also test for treatment heterogeneity by whether workers are currently employed outside of Prolific, a proxy for their outside options. Perhaps surprisingly, there is no differential effect on retention, but the negative effect on performance is driven by workers who are currently employed outside Prolific (Appendix Tables F3 and F4), so perhaps workers with higher outside options are more likely to mentally disengage when they perceive discrimination. This dimension of heterogeneity was not pre-registered.

Future work. Appendix Table F5 combines all pre-registered dimensions of heterogeneity; the effect on willingness to pay to be able to choose one's own manager is largest for workers who reported experiencing discrimination at work in the past. There is no differential effect by whether workers are currently employed outside Prolific or not (Appendix Table F6).

Table F1: Effects on retention in the manager sample of the promotion experiment, all measures of heterogeneity

	Num paragraphs	More than 6	Did all 18
	(1)	(2)	(3)
Non-blind manager arm	-0.916*	-0.034	-0.107**
	(0.480)	(0.034)	(0.048)
Non-blind manager arm*Minority Men	1.429**	0.100**	0.112*
	(0.614)	(0.046)	(0.060)
Non-blind manager arm*Exp Past Discr	-0.545	0.004	-0.066
	(0.488)	(0.034)	(0.049)
Non-blind manager arm*High Discr Beliefs	0.258	-0.011	0.038
	(0.483)	(0.033)	(0.049)
Non-blind manager arm*High Confidence	0.610	0.003	0.108**
	(0.481)	(0.033)	(0.048)
N	1293	1293	1293
Control Mean	15.916	0.915	0.776

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, allowing the effect to vary by all pre-registered dimensions of experiment-specific heterogeneity. The outcomes and sample are the same as in Table 2, but white men are additionally dropped from the sample. The specification is also the same, but the treatment indicator is also interacted with being a racial minority man (omitted group: women), reporting experiencing discrimination in the past (omitted group: not), having above-median beliefs about the prevalence of past experienced discrimination in this sample (omitted group: below-median), and having above-median confidence in the spelling, science, and grammar quizzes in the baseline survey (omitted group: below-median). Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table F2: Effects on effort in the manager sample of the promotion experiment, all measures of heterogeneity

	Minutes	Clicks	Correct	Incorrect	Bonus
	(1)	(2)	(3)	(4)	(5)
Non-blind manager arm	-0.056	2.656	0.070	0.583*	-0.015
	(0.104)	(3.119)	(0.441)	(0.335)	(0.040)
Non-blind manager arm*Minority Men	0.155	-6.437*	1.136**	-0.284	0.079
	(0.131)	(3.295)	(0.539)	(0.534)	(0.051)
Non-blind manager arm*Exp Past Discr	0.111	1.589	0.436	0.189	0.012
	(0.109)	(3.606)	(0.448)	(0.394)	(0.043)
Non-blind manager arm*High Discr Beliefs	-0.087	-1.102	-0.396	-0.450	0.014
	(0.109)	(3.447)	(0.455)	(0.418)	(0.043)
Non-blind manager arm*High Confidence	-0.006	-1.691	0.167	-0.377	-0.002
	(0.106)	(3.479)	(0.442)	(0.384)	(0.042)
N	1295	1295	1295	1295	1295
Control Mean	4.957	36.804	14.817	3.422	0.919

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, allowing the effect to vary by all pre-registered dimensions of experiment-specific heterogeneity. The outcomes and sample are the same as in Table 3, but white men are additionally dropped from the sample. The specification is the same as in Appendix Table F1. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table F3: Effects on retention in the manager sample of the promotion experiment, heterogeneity by employment outside Prolific

	Num paragraphs	More than 6	Did all 18
	(1)	(2)	(3)
Non-blind manager arm*Employed	-0.574*	-0.015	-0.060**
	(0.296)	(0.021)	(0.030)
Non-blind manager arm*Not employed	-0.355	-0.017	-0.028
	(0.410)	(0.028)	(0.041)
p-values:			
Employed = not employed	*0.649*	*0.952*	*0.507*
N	1387	1387	1387
Control Mean, employed	15.927	0.914	0.780
Control Mean, not employed	15.895	0.916	0.768

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 workers who are currently employed outside prolific and those who are not. The outcomes and sample are the same as in Table 2. The specification is the same, but the indicator for treatment is replaced with two interactions between treatment and indicators for being employed or not being employed. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table F4: Effects on effort and performance in the manager sample of the promotion experiment, heterogeneity by employment outside Prolific

	Minutes	Clicks	Correct	Incorrect	Bonus
	(1)	(2)	(3)	(4)	(5)
Non-blind manager arm*Employed	-0.001	0.809	0.523*	0.149	0.014
	(0.067)	(2.221)	(0.277)	(0.262)	(0.026)
Non-blind manager arm*Not employed	-0.076	0.312	-0.063	0.370	-0.005
	(0.087)	(3.074)	(0.372)	(0.319)	(0.035)
p-values:					
Employed = not employed	*0.475*	*0.896*	*0.196*	*0.564*	*0.661*
N	1389	1389	1389	1389	1389
Control Mean, employed	4.909	37.302	14.807	3.635	0.925
Control Mean, not employed	5.050	35.844	14.835	3.013	0.909

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 workers who are currently employed outside prolific and those who are not. The outcomes and sample are the same as in Table 3. The specification is the same as Appendix Table F3. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table F5: Effects on future labor supply in the manager sample of the promotion experiment, all measures of heterogeneity

	RW (Same)	WTP cutoff	Coop. task RW	WTP new manager
	(1)	(2)	(3)	(4)
Non-blind manager arm	0.017	0.015	0.005	0.064
	(0.019)	(0.009)	(0.028)	(0.041)
Non-blind manager arm*Minority Men	0.006	0.002	-0.027	-0.074
	(0.025)	(0.013)	(0.035)	(0.051)
Non-blind manager arm*Exp Past Discr	0.018	0.003	0.004	0.064*
	(0.020)	(0.009)	(0.028)	(0.039)
Non-blind manager arm*High Discr Beliefs	0.009	-0.003	0.002	-0.052
	(0.020)	(0.010)	(0.028)	(0.038)
Non-blind manager arm*High Confidence	-0.019	-0.007	0.007	0.007
	(0.020)	(0.009)	(0.028)	(0.039)
N	1247	1235	1256	957
Control Mean	0.265	-0.014	0.341	0.223

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, allowing the effect to vary by all pre-registered dimensions of experiment-specific heterogeneity. The outcomes and sample are the same as in Table 4, but white men are additionally dropped from the sample. The specification is the same as in Appendix Table F1. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

Table F6: Effects on future labor supply in the manager sample of the promotion experiment, heterogeneity by employment outside Prolific

	RW (Same)	WTP cutoff	Coop. task RW	WTP new manager
	(1)	(2)	(3)	(4)
Non-blind manager arm*Employed	0.026**	0.012**	0.017	0.046*
	(0.012)	(0.006)	(0.017)	(0.025)
Non-blind manager arm*Not employed	0.018	0.007	-0.009	0.032
	(0.016)	(0.008)	(0.024)	(0.031)
p-values:				
Employed = not employed	*0.711*	*0.601*	*0.351*	*0.727*
N	1338	1325	1349	1030
Control Mean, employed	0.269	-0.016	0.337	0.228
Control Mean, not employed	0.257	-0.010	0.347	0.214

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 workers who are currently employed outside prolific and those who are not. The outcomes and sample are the same as in Table 4. The specification is the same as Appendix Table F3. Standard errors, in parentheses, are robust to heteroskedasticity. Significance at the 0.1, 0.05, and 0.01 levels indicated by *, **, and ***, respectively.

G Robustness analyses

G.1 Alternative inference methods

Randomization inference. Randomization inference, or approximate permutation testing, 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 β^k using the re-assigned "treatment" with the same specification as the main analysis. The distribution of β^k's provides the empirical distribution of the treatment effect parameter under the null hypothesis that $\beta^k = 0$. The two-sided randomization inference p-value is the fraction of absolute values of β^k's that are greater than the absolute value of the observed β^{true} estimated using the real treatment assignment.

All of the main results and conclusions of the paper are unaffected by using randomization inference. Appendix Figure G1 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 0.06. The p-value on whether workers share with their manager (originally significant at the 10 percent level) is only marginally significant with a p-value of 0.16 when using randomization inference. Appendix Figure G2 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. The effects of increasing minority-group representation retain p-values less than 0.02 (Appendix Figure G3). The effects of learning that the demographic-blind algorithm previously promoted mostly white men on performance retain significance at the 10 percent level (Appendix Figure G4).

Multiple hypothesis testing. All of the figures reference above also plot the corresponding asymptotic p-values when they are corrected for multiple hypothesis testing; all results except the effects of perceived algorithmic discrimination are unaffected. Specifically, I control the family-wise error rate using the Romano-Wolf (2005; 2016) correction with 500 bootstrap iterations. In Appendix Figure G1 the four "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. The effects of being in the non-blind manager arm relative to the demographic-blind manager arm are generally robust; p-values change the most for the future labor supply outcomes, which in some cases are now only marginally significant (p-values around 0.16). Again, 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 (Appendix Figure G2). Appendix Figure G3 does not include MHT-corrected p-values 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. The 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.

G.2 Attrition and attention

Differential attrition and attention also cannot explain any of the results Appendix Table G1, 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 avatars, race, and gender (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 G21 shows that there is similarly no differential attrition in the hiring experiment.

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 G1, Panel D, column 3). Note that workers who failed the attention check are included in the main sample, and the results are robust to dropping them, as discussed in the next section. Section 2.4 discusses overall experimental fidelity and shows that worker attention to and comprehension of the experiment was high overall.

G.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. Appendix Figures G5-G12 show the resulting estimates for the effects of being in the non-blind manager arm relative to the blind manager arm in the promotion experiment on rates of perceived discrimination and all of the main worker-behavior outcomes. Appendix Figures G13 and G14 test the robustness of the effects of anti-bias hiring policies on rates of perceived discrimination, Appendix Figure G15 tests the robustness of the effects of minority-group representation, and G16-G20 test the robustness of the effects of perceived algorithmic discrimination.

G.3.1 Sample definitions

The main results restrict the sample to workers who were not selected by any of the procedures, as preregistered. 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 specification charts iterate over other sample definitions: further restricting to workers who passed the "attention check" in the final survey section or expanding to the full sample of workers who completed the experimental survey. The effects of perceived manager discrimination on behavior, effects of anti-bias hiring procedures on perceived discrimination, and effects of minority-group representation on perceived discrimination are generally unchanged by either of these changes. The effects of perceived algorithmic discrimination are again 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 the sample that also passed the attention check (Appendix Figure G19). Being in the non-blind manager arm also seems to negatively affect time spent per paragraph when restricting to workers who passed the attention check (Appendix Figure G7).

G.3.2 Included controls

The main specification controls for demographic information (age, education, income, race, gender, family status), 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, the double-post-lasso procedure was the following: 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).[44] 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.

G.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 bias from experimenter demand effects, relative to an older literature of lab experiments in psychology in which the phenomena was originally documented.[45]

[44] 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, indicators for being currently or ever employed outside of Prolific, 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, and scales of racial and gender identity centrality. 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.

[45] Mummolo and Peterson (2019) show that giving participants information about experimenter hypotheses does not affect partici-

Unsurprisingly, Appendix Table G2 (promotion experiment) and Appendix Figure G22 (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.[46] In the promotion experiment 0-1 percent and 3-4 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 0-4 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 5pp (se=4pp) and 10pp (se=2pp) 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 5-9pp (depending on when they were ask 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 2-9pp more likely to do so.

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 G22). 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 differential 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.[47] Appendix Figures G23 and G24 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 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,

pant behavior or estimated treatment effects, even when participants are financially incentivized to conform to those hypotheses and participants are confirmed to be 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.

[46] 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. See Appendix H for details. In the promotion experiment, these were coded by external coders. Due to budget and time constraints, in the hiring experiment, the coded variables were generated from the text responses by this author.

[47] de Quidt et al. (2018) show that a strong message telling workers to worker 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.

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 Figures G25 and G26 shows that the effects of perceived algorithmic discrimination on performance are also robust to this adjustment.

Appendix Figures G27 show 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 (but still statistically significant) 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 25 percent rather than about 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 effect of blinding managers and algorithms to demographics are unchanged. The lower bound on the effects of using an algorithm rather than a manager fall to about half the magnitude and are no longer statistically significant (Appendix Figure G28).

G.5 Multiple price list elicitations

The four measures of future labor supply were elicited using multiple price lists (MPL).[48] These were reservation wages for future work when (1) one 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. Appendix D shows the exact implementation of each MPL.

Following the literature, these were incentivized by telling workers that I would randomly choose one row of the MPL and implement the choice of the worker in that row for randomly-selected workers. That said, workers could report low reservation wages and simply 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, report downward-sloping labor supply or upward-sloping demand for choosing their own manager). Appendix Table G3 compares the workers who are dropped for this reason with other workers. Panel A shows that these workers are less attentive and perform worse on the proofreading task than those who answer the MPL questions sensibly, but have similar education levels. 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 in the dictator game. In other words, column (2) of Panel B shows that the main results for worker behavior outcomes are unchanged when restricting the sample to those for whom I observe all reservation wage and willingness to pay outcomes.

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

G.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 A9). 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 in the paper despite not being pre-registered. There were also differences across arms in what fraction of workers who saw three white men; Appendix Figure 5 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. Thus, a much larger fraction of the sample is white women than planned. That said, I am still able to test for racial and gender heterogeneity and obtain moderately precise results.

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 in the promotion experiment, and one evaluation group of 360 in the 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.

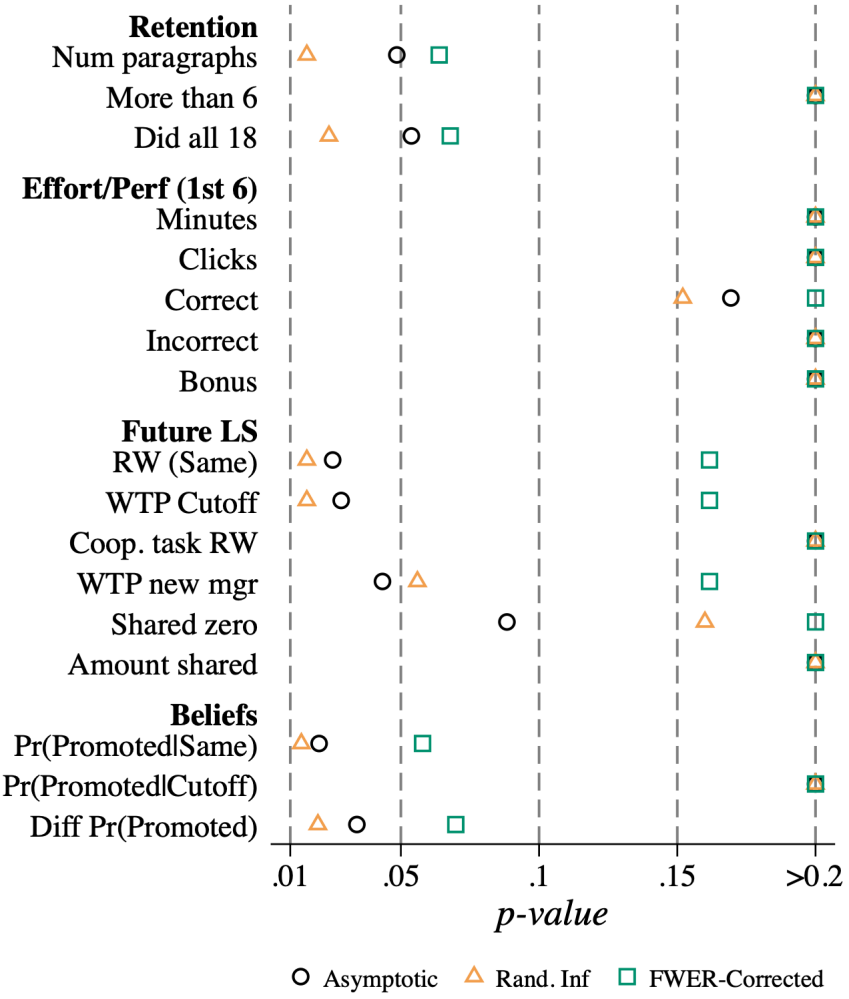
Added controls for previously selected workers' education and baseline performance. In the control groups, workers saw the education and baseline performance of the workers that their manager or the algorithm had selected in the historical sample. Comparing the treated groups, which also saw previously 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 G.3 however, this does not affect the main conclusions. Estimates without these controls are in Appendix Figures G13-G20.

De-emphasis of IV specification. Section 3.1 and Appendix E discusses the challenges associated with interpreting the IV specification, despite it being pre-registered, and the appendix presents all of the results for the pre-registered specifications.

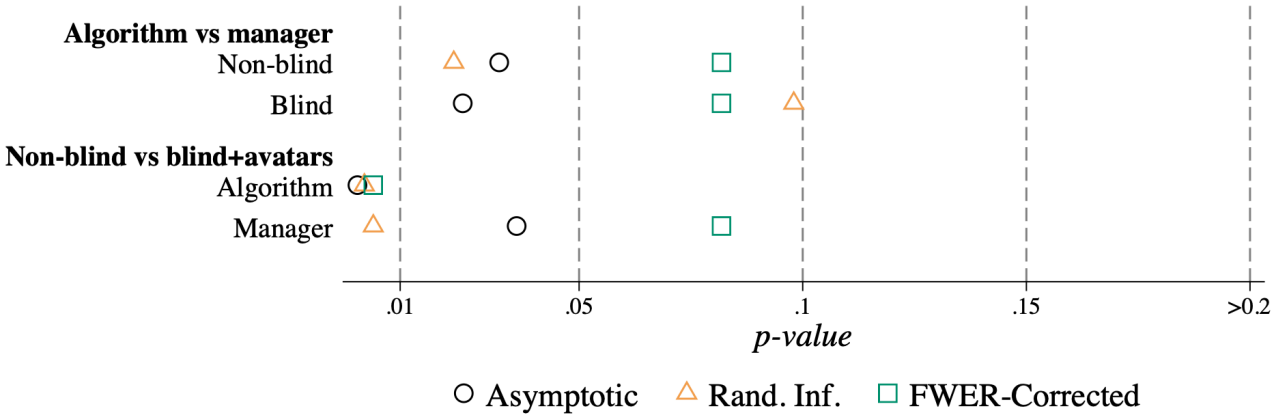
Implicit measures of perceived discrimination. The paper focuses on one primary measure of perceived discrimination and shows results for the other pre-registered measures of explicit perceived discrimination, but does not show results for the pre-registered measures of *implicit* perceived discrimination. 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 generally imprecise. Due to space constraints, these are available upon request.

Figure G1: Randomization inference and multiple hypothesis testing (promotion experiment, manager arms)



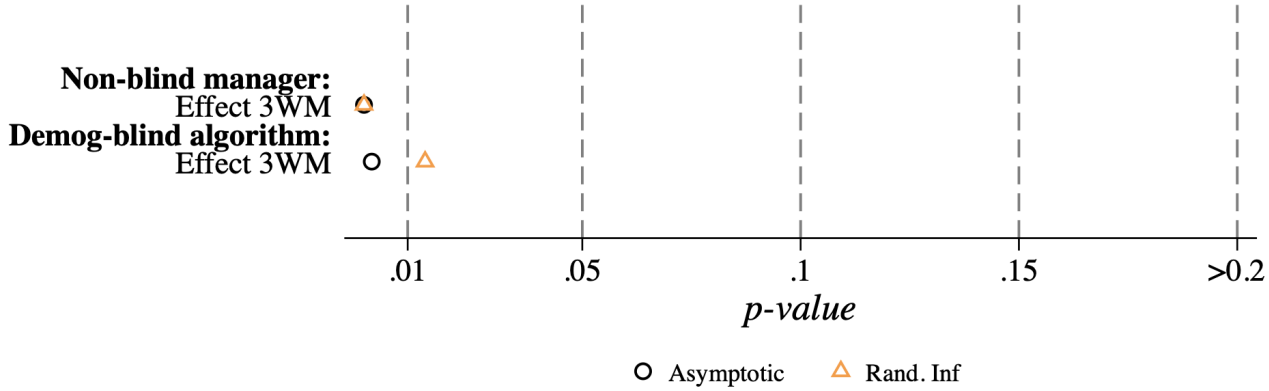
Note: This figure plots the *p*-values for the effects of being in the non-blind manager group relative to the demographic-blind manager group in the promotion experiment on the key outcomes of interest for the specifications in Tables 2, 3, 4, 5, and 6, with the *p*-values calculated asymptotically (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 G2: Randomization inference and multiple hypothesis testing (hiring experiment)



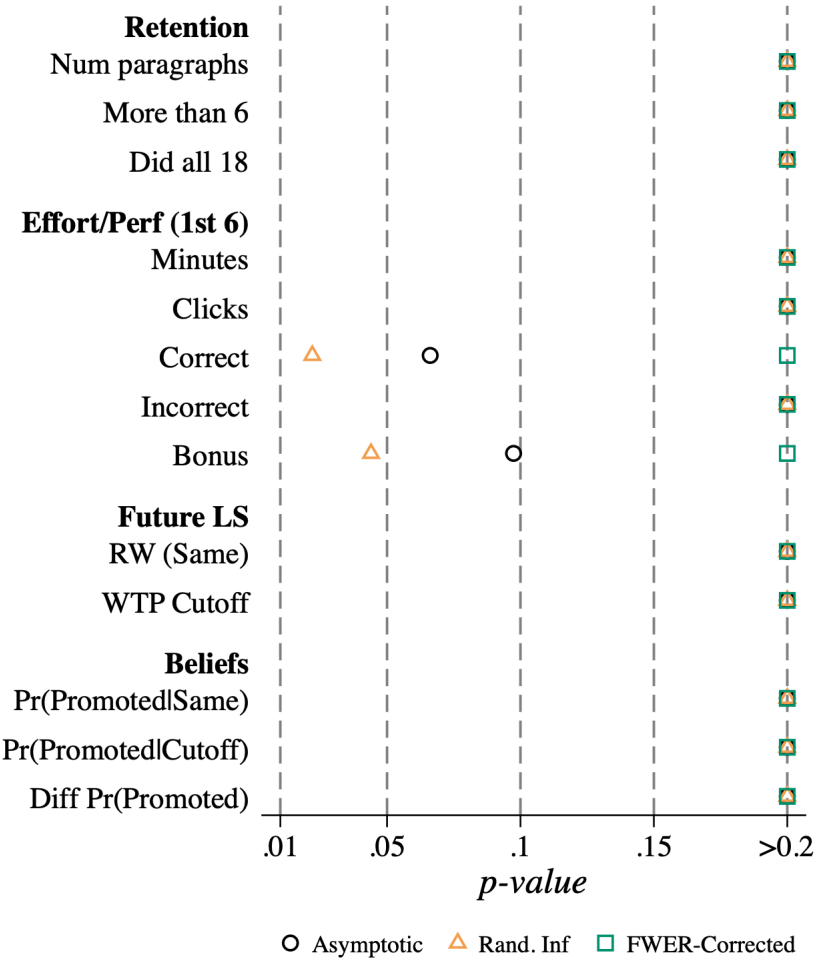
Note: This figure plots the *p*-values for the differences in perceived discrimination between treatment arms in Figure 4, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity), with randomization inference, and corrected for multiple hypothesis testing using the Romano-Wolf correction.

Figure G3: Randomization inference and multiple hypothesis testing (historical representation)



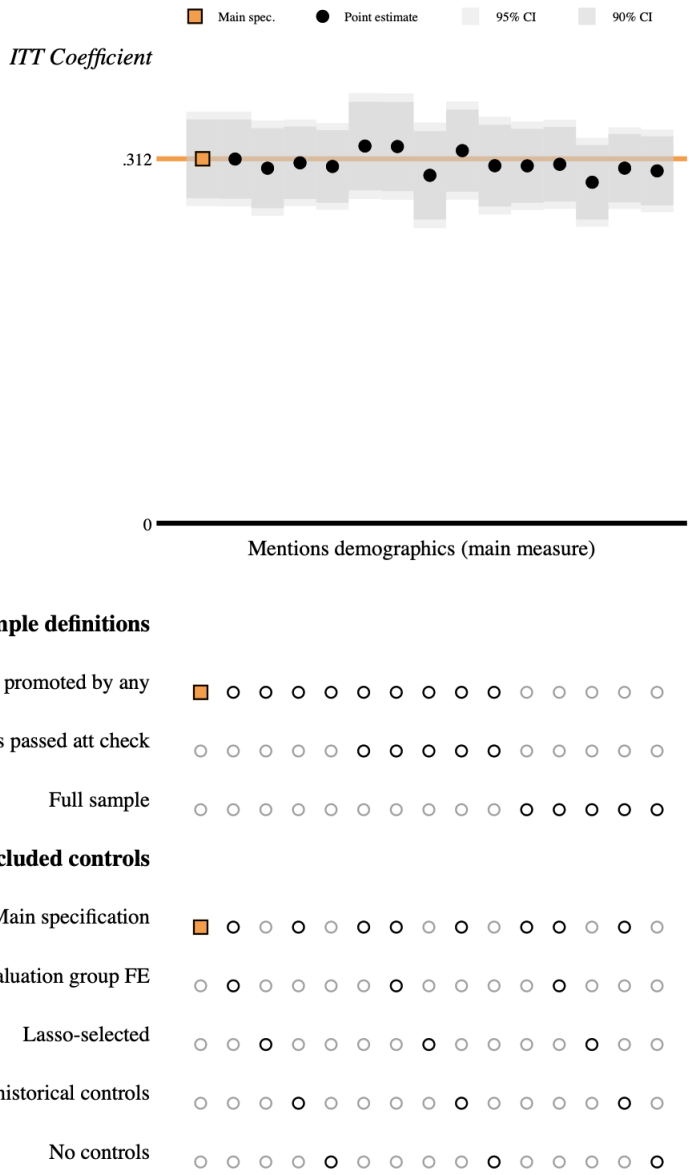
Note: This figure plots the *p*-values for the differences in perceived discrimination between treatment arms in Figure 4, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity) and with randomization inference.

Figure G4: Randomization inference and multiple hypothesis testing (promotion experiment, algorithm sub-arms)



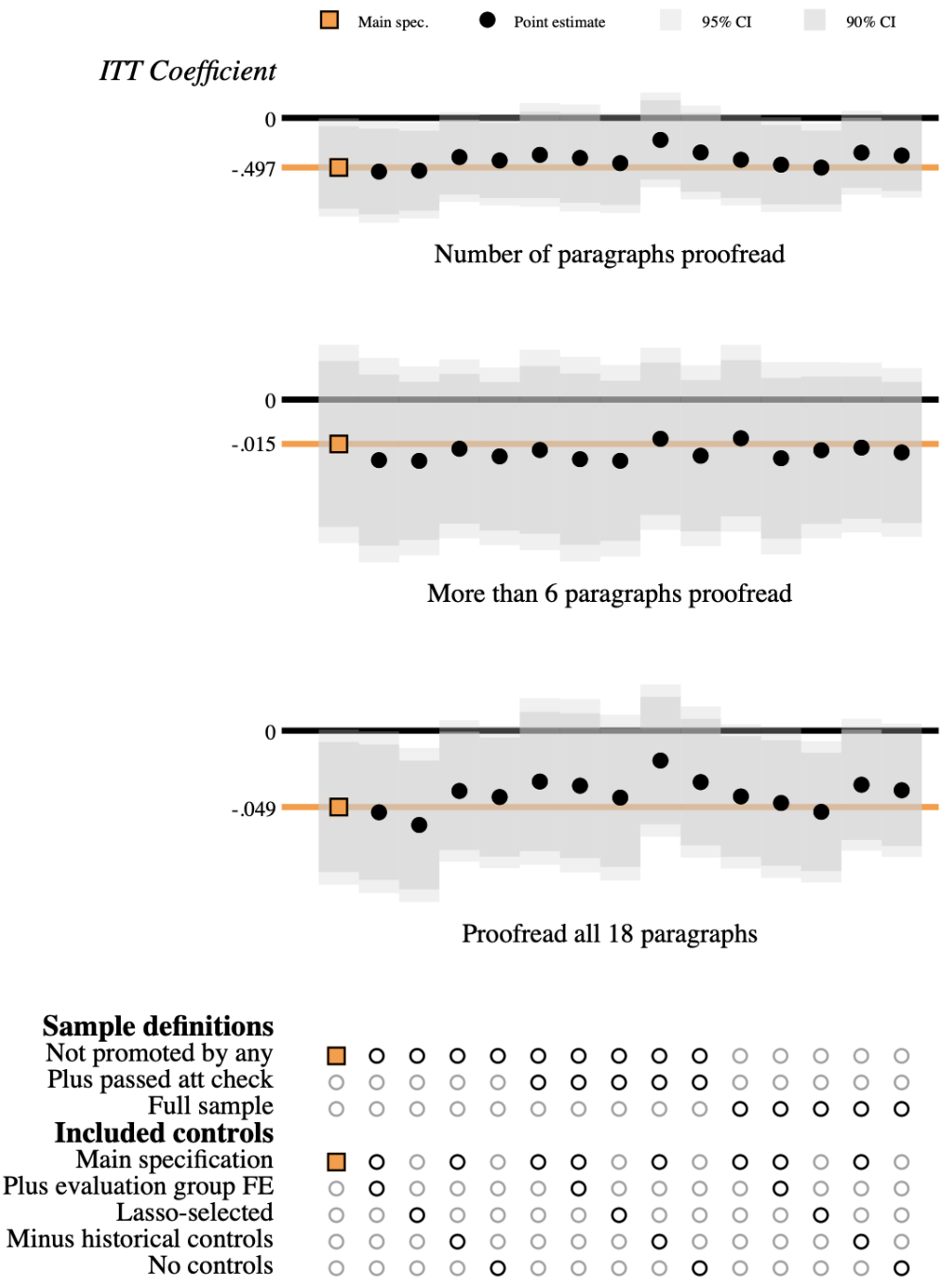
Note: This figure plots the *p*-values for the effects of learning that the demographic-blind algorithm previously promoted mostly white men in the promotion experiment on the key outcomes of interest for the specifications in Figure 7, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity), with randomization inference, and corrected for multiple hypothesis testing using the Romano-Wolf correction.

Figure G5: Alternative specifications: Effect of being in non-blind manager arm on perceived discrimination



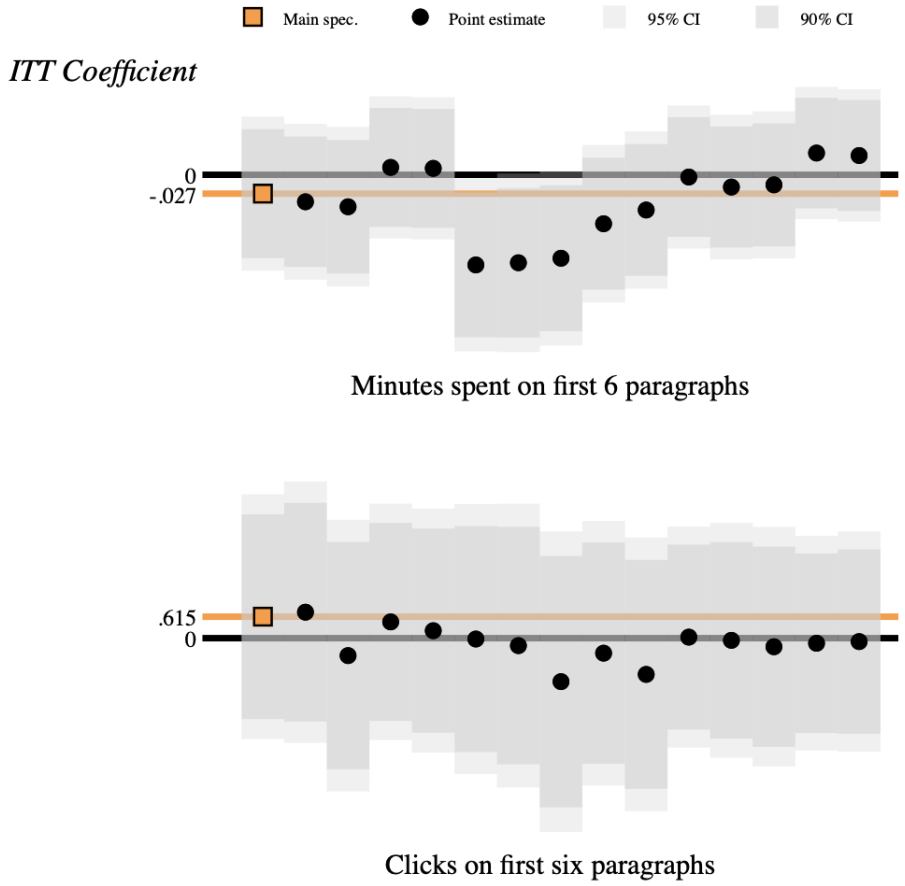
Note: This figure plots the treatment effect of being in the non-blind manager group on the main measure of perceived discrimination in the promotion experiment, relative to the demographic-blind manager group. The figure iterates over sample restrictions and choices of control variables, starting with the main specification. Starting with the main sample definition (workers who would not have been promoted 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 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 G6: Alternative specifications: Effect of being in non-blind manager arm on retention



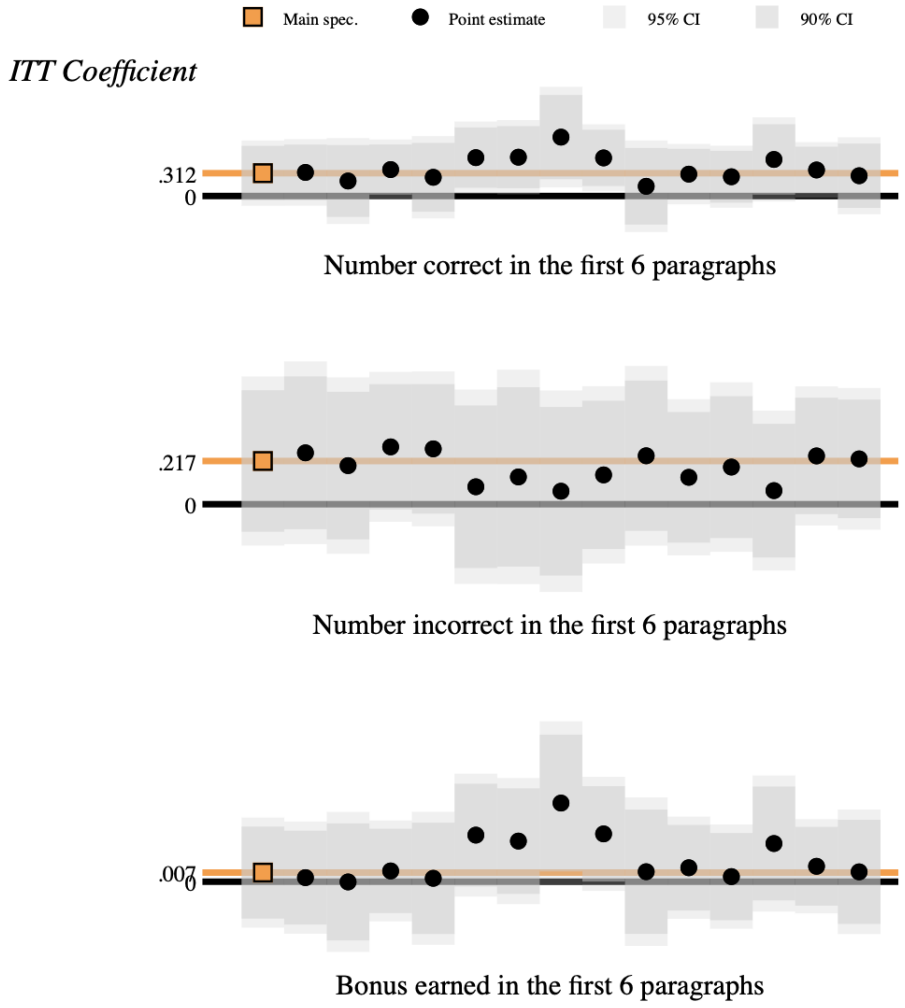
Note: This figure plots the treatment effect of being in the non-blind manager group on retention in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 2. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G7: Alternative specifications: Effect of being in non-blind manager arm on effort



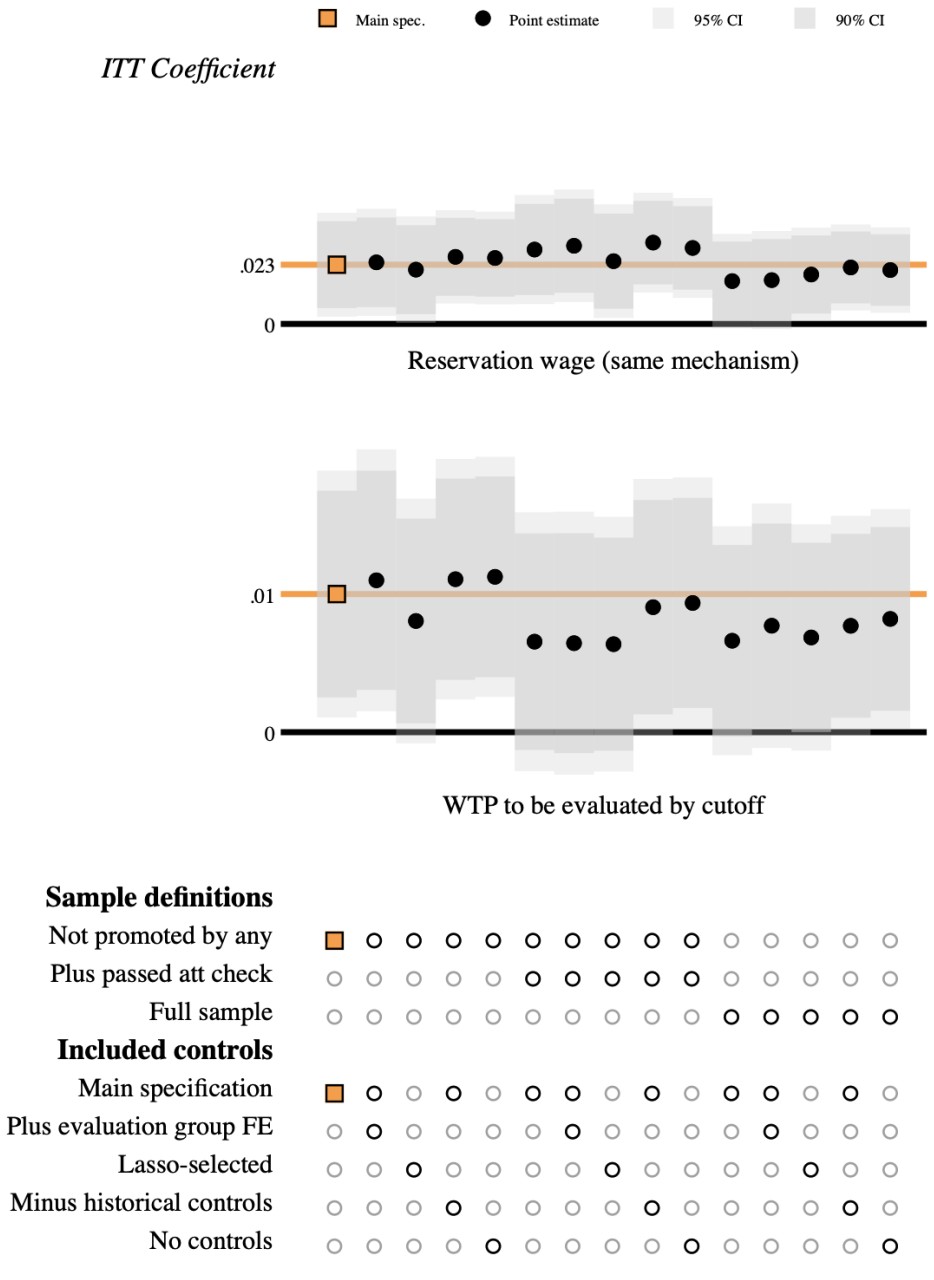
Note: This figure plots the treatment effect of being in the non-blind manager group on effort in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in columns 1 and 2 of Table 3. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G8: Alternative specifications: Effect of being in non-blind manager arm on performance



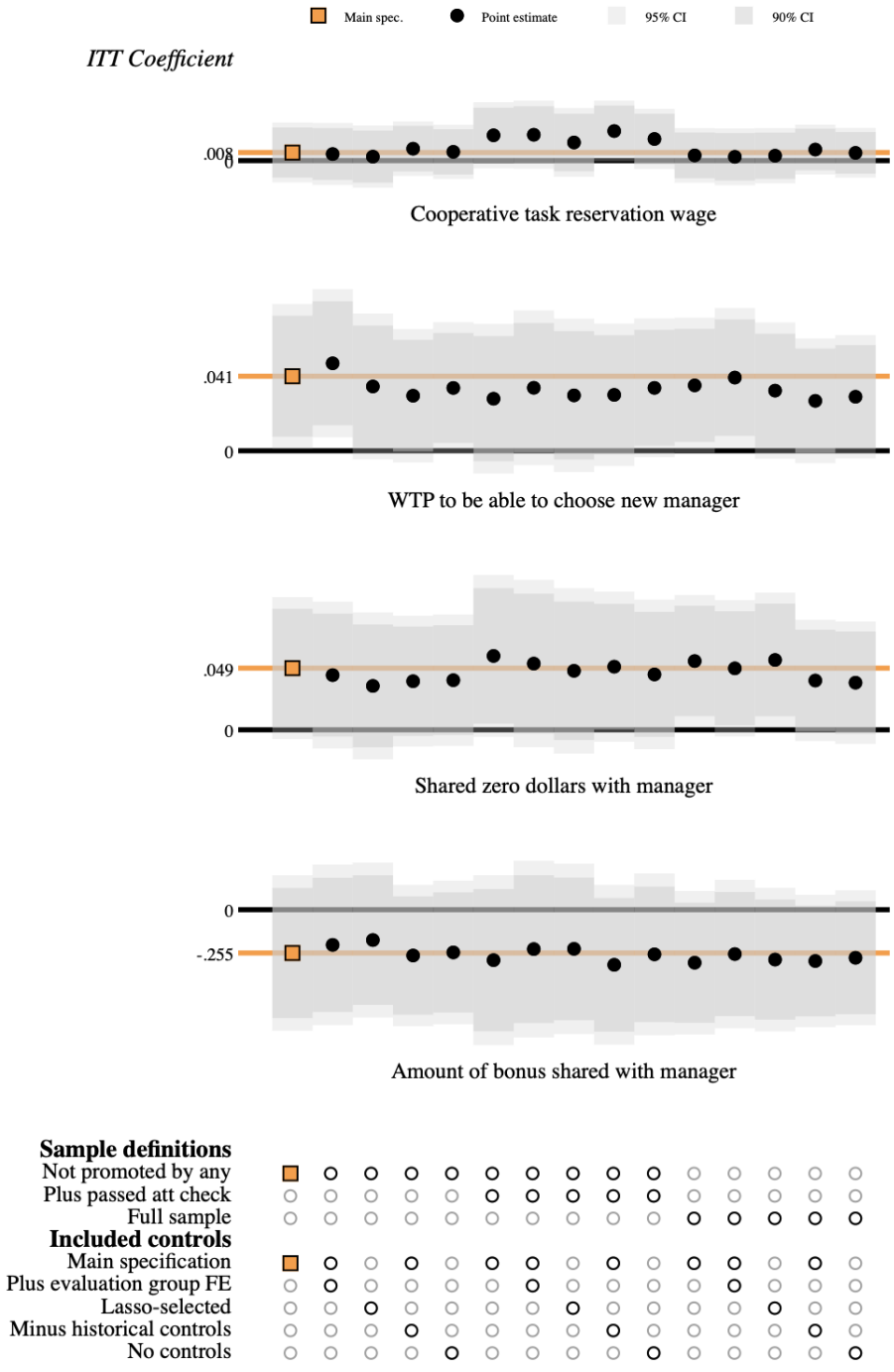
Note: This figure plots the treatment effect of being in the non-blind manager group on performance in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in columns 3-5 of Table 3. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G9: Alternative specifications: Effect of being in non-blind manager arm on future labor supply for the same job



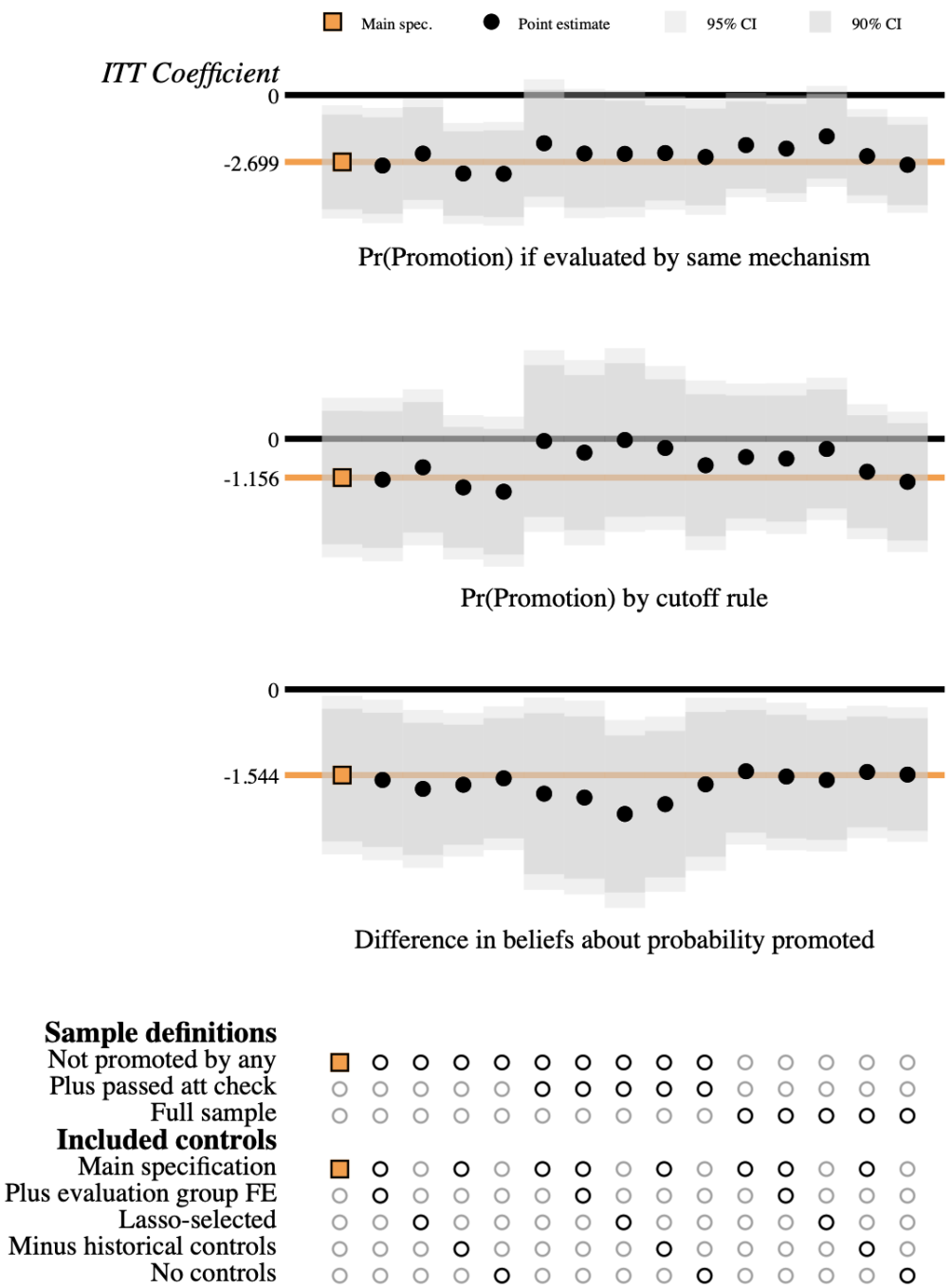
Note: This figure plots the treatment effect of being in the non-blind manager group on reservation wages in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 4. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G10: Alternative specifications: Effect of being in non-blind manager arm on future labor supply (cooperative task) and generosity



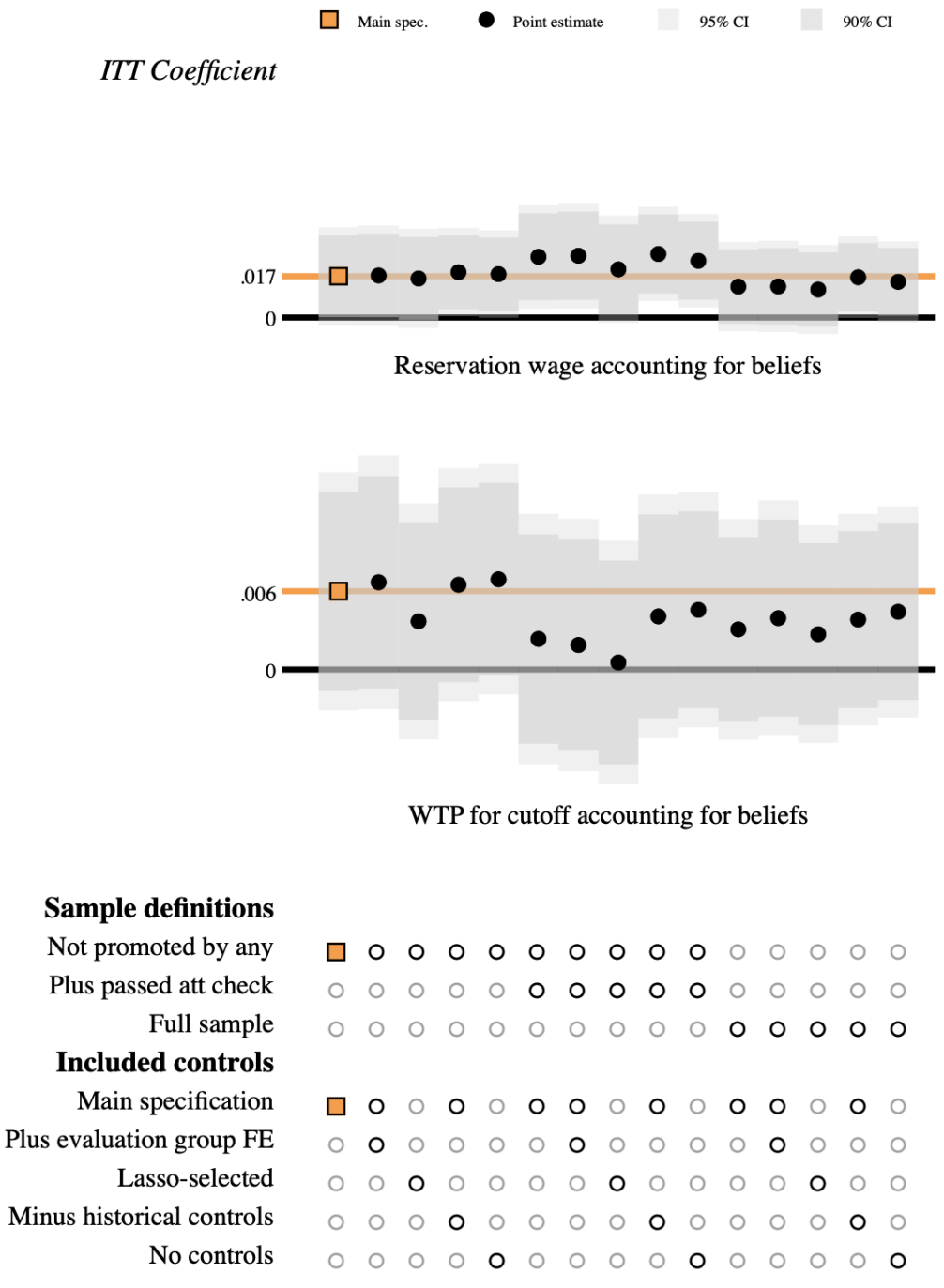
Note: This figure plots the treatment effect of being in the non-blind manager group on willingness to interact and share with managers in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 5. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G11: Alternative specifications: Effect of being in non-blind manager arm on beliefs about promotion



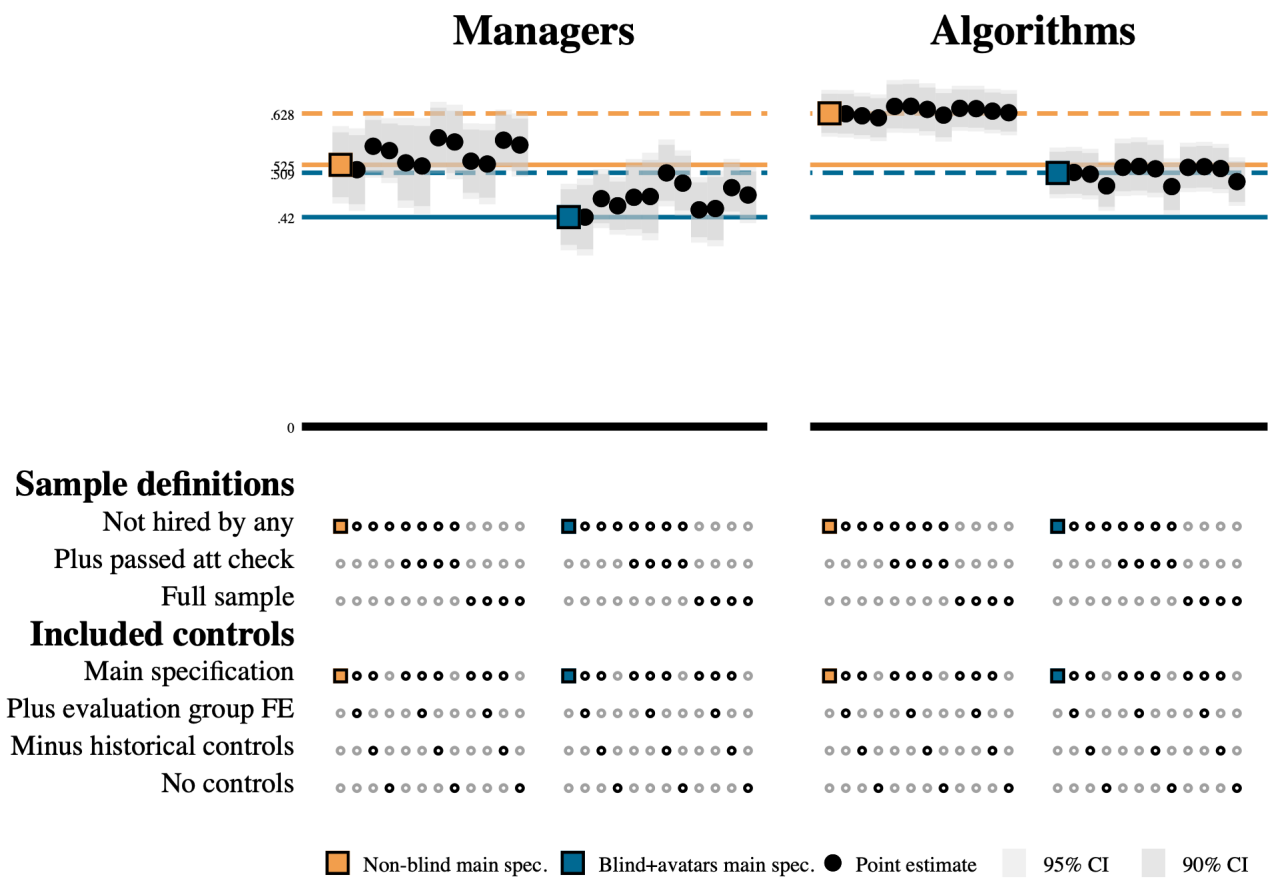
Note: This figure plots the treatment effect of being in the non-blind manager group on beliefs about promotion in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 6. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G12: Alternative specifications: Effect of being in non-blind manager arm on future labor supply, accounting for beliefs



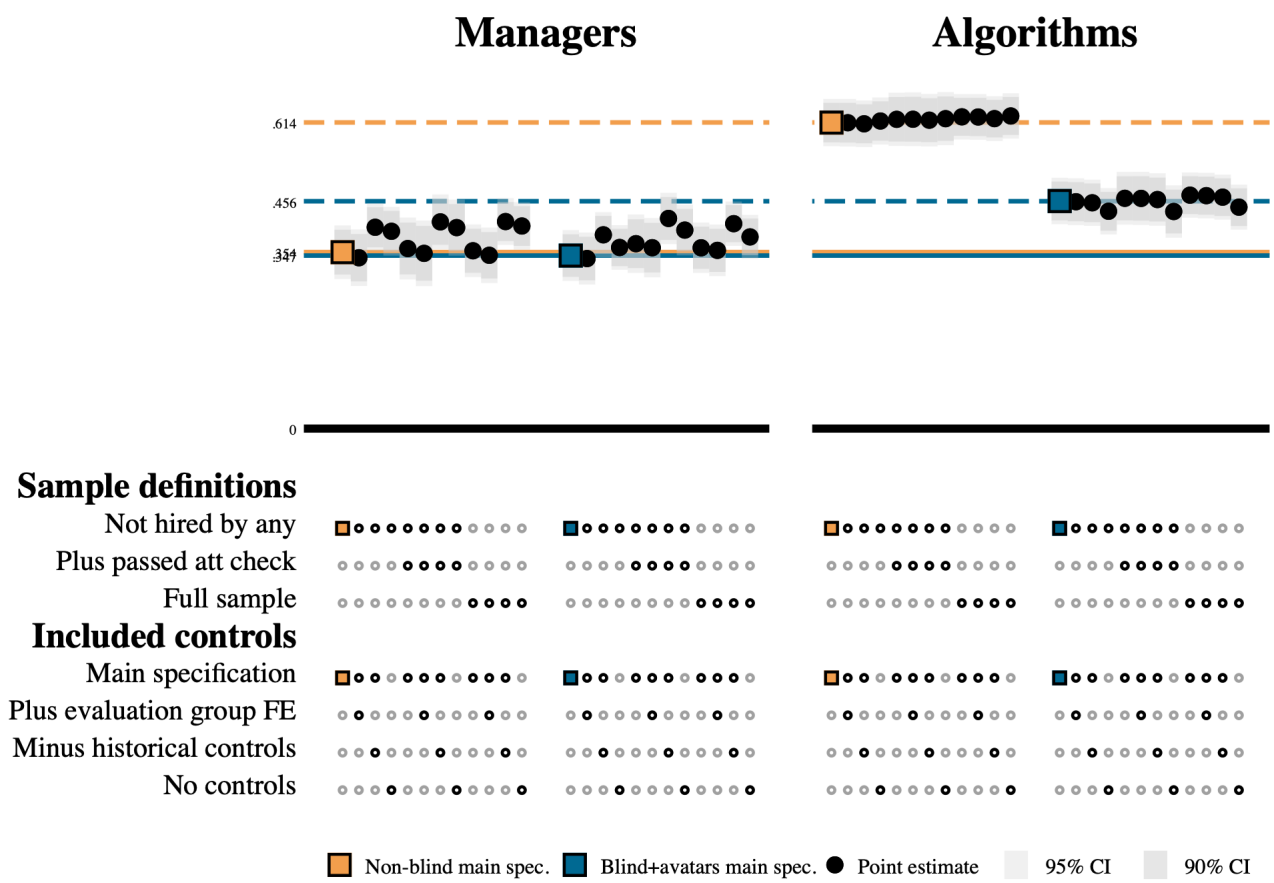
Note: This figure plots the treatment effect of being in the non-blind manager group on reservation wages after accounting for the effects on beliefs about promotion in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in the second specification (in blue) in Figure 3. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G13: Alternative specifications: Hiring experiment (if previous hires are all white men)



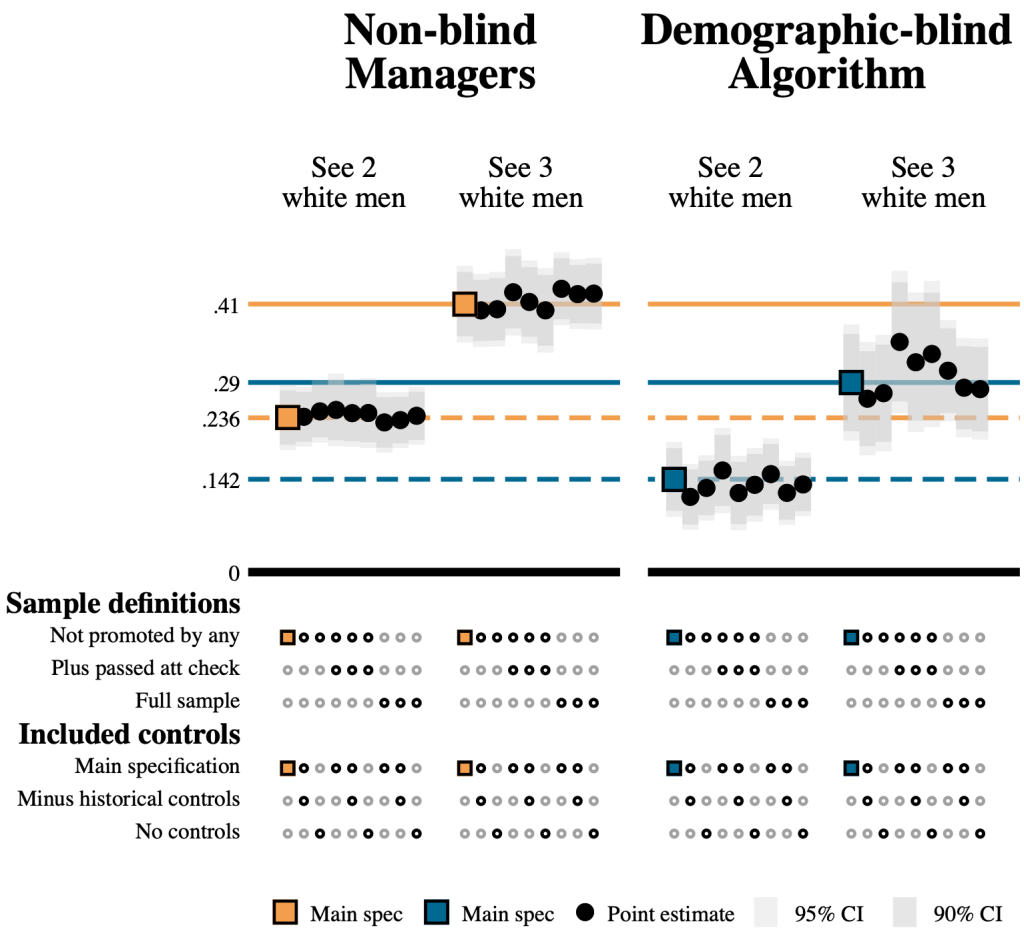
Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the hiring experiment. It parallels Figure 4, with the sample restricted to workers who saw only white men being previously hired. The figure iterates over other sample restrictions and choices of 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 G14: Alternative specifications: Hiring experiment (regardless of demographic makeup of previous hires)



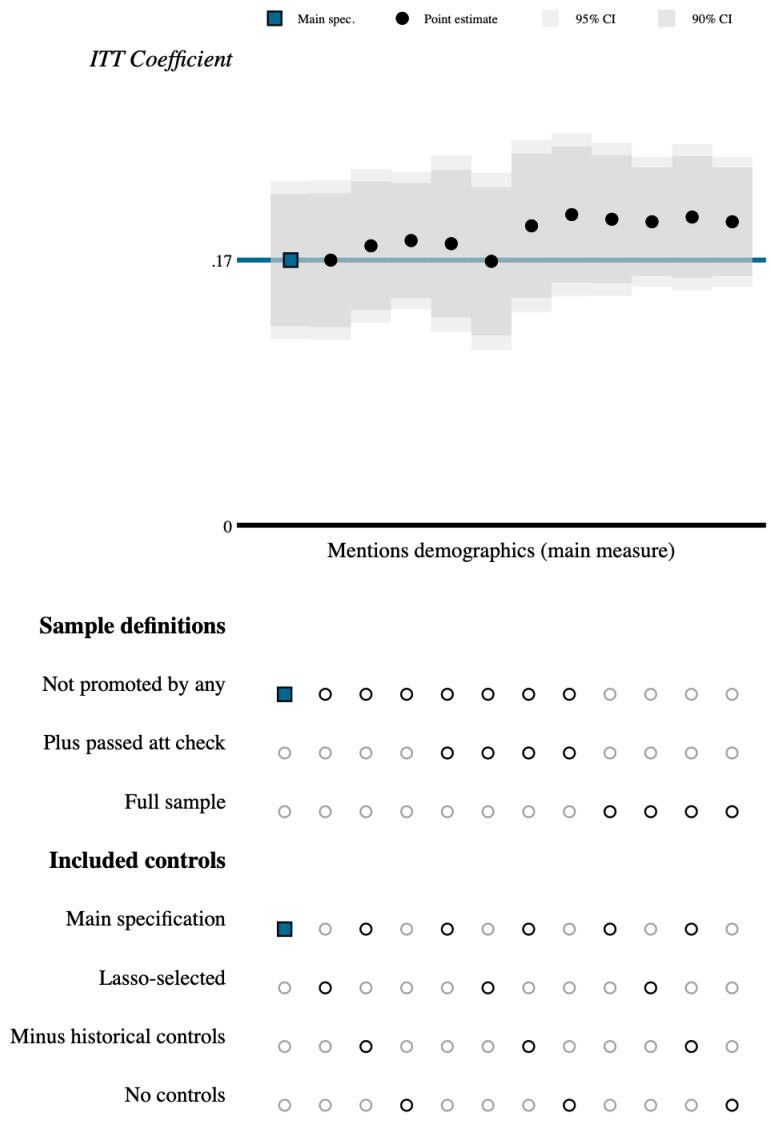
Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the hiring experiment, paralleling Appendix Figure A23 in using the full sample (not accounting for differences in the share of previous hires who are white men). Otherwise, it is analogous to Appendix Figure G13.

Figure G15: Alternative specifications: Effects of minority-group representation



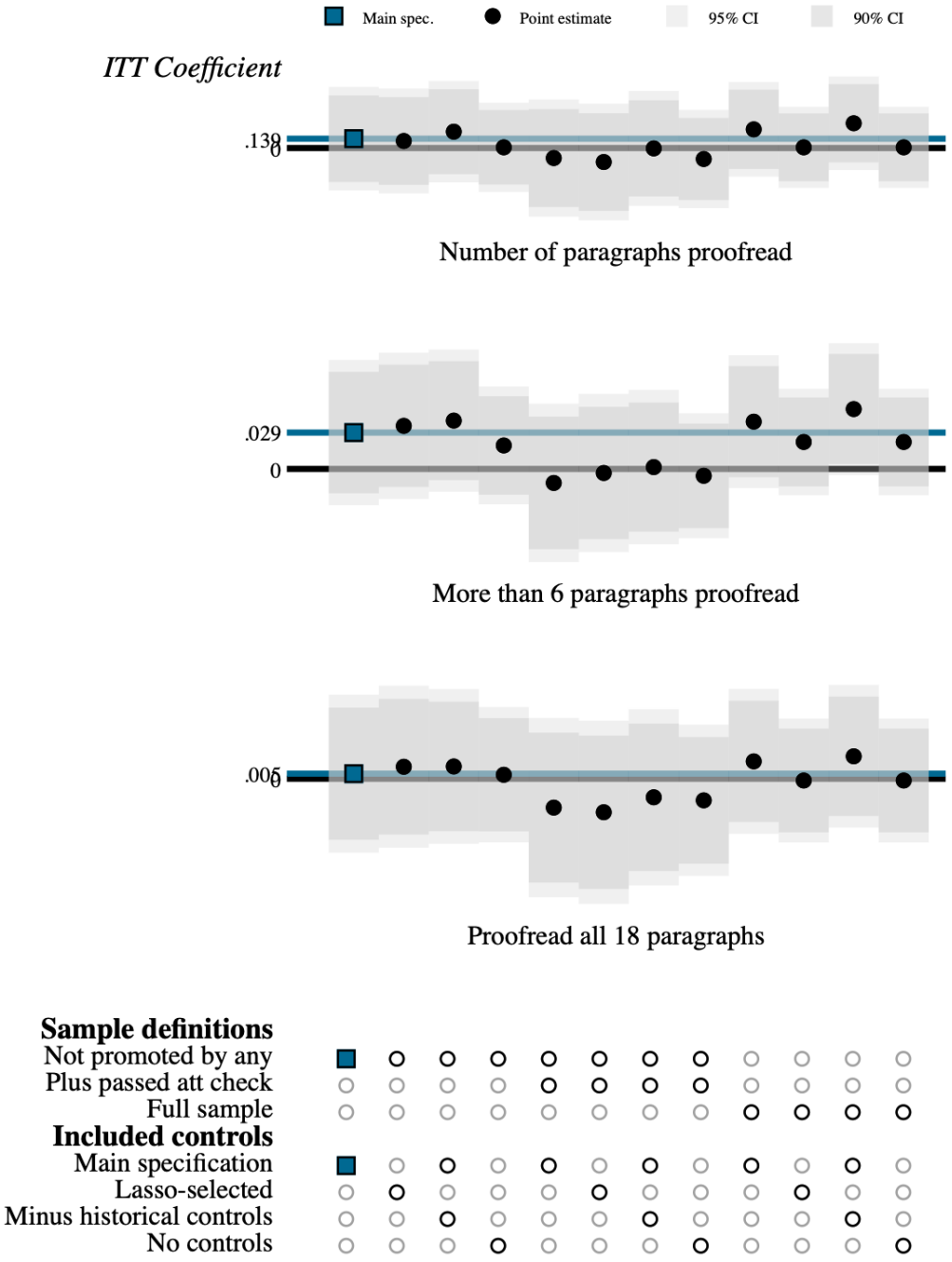
Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the promotion experiment. It parallels Figure 5, plotting only the share perceiving discrimination in the non-blind manager arm and the demographic-blind algorithm arm in which workers saw previously-promoted workers' demographics (though these rates are estimated using the same specification as Figure 5. 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 G16: Alternative specifications: Effects of seeing avatars on perceived algorithmic discrimination (promotion experiment)



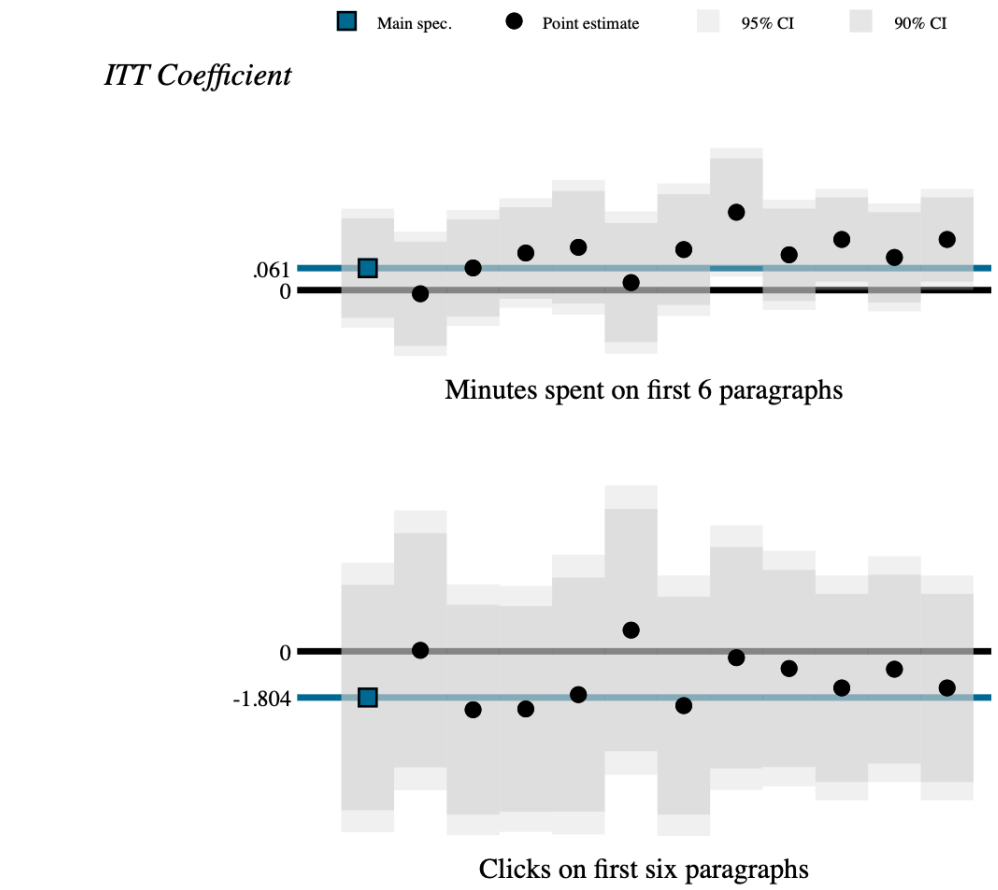
Note: This figure plots the treatment effect of seeing that the demographic-blind algorithm previously promoted mostly white men on perceived discrimination in the promotion experiment, relative to the other workers in the demographic-blind algorithm group who do not see the avatars, race, and gender of previously-promoted workers. Because the randomization within the algorithm arm was across rather than within evaluation groups, this chart does not include a specification that adds evaluation group fixed effects. Otherwise, the figure is analogous to Appendix Figure G5.

Figure G17: Alternative specifications: Effects of seeing avatars in the algorithm arm on retention



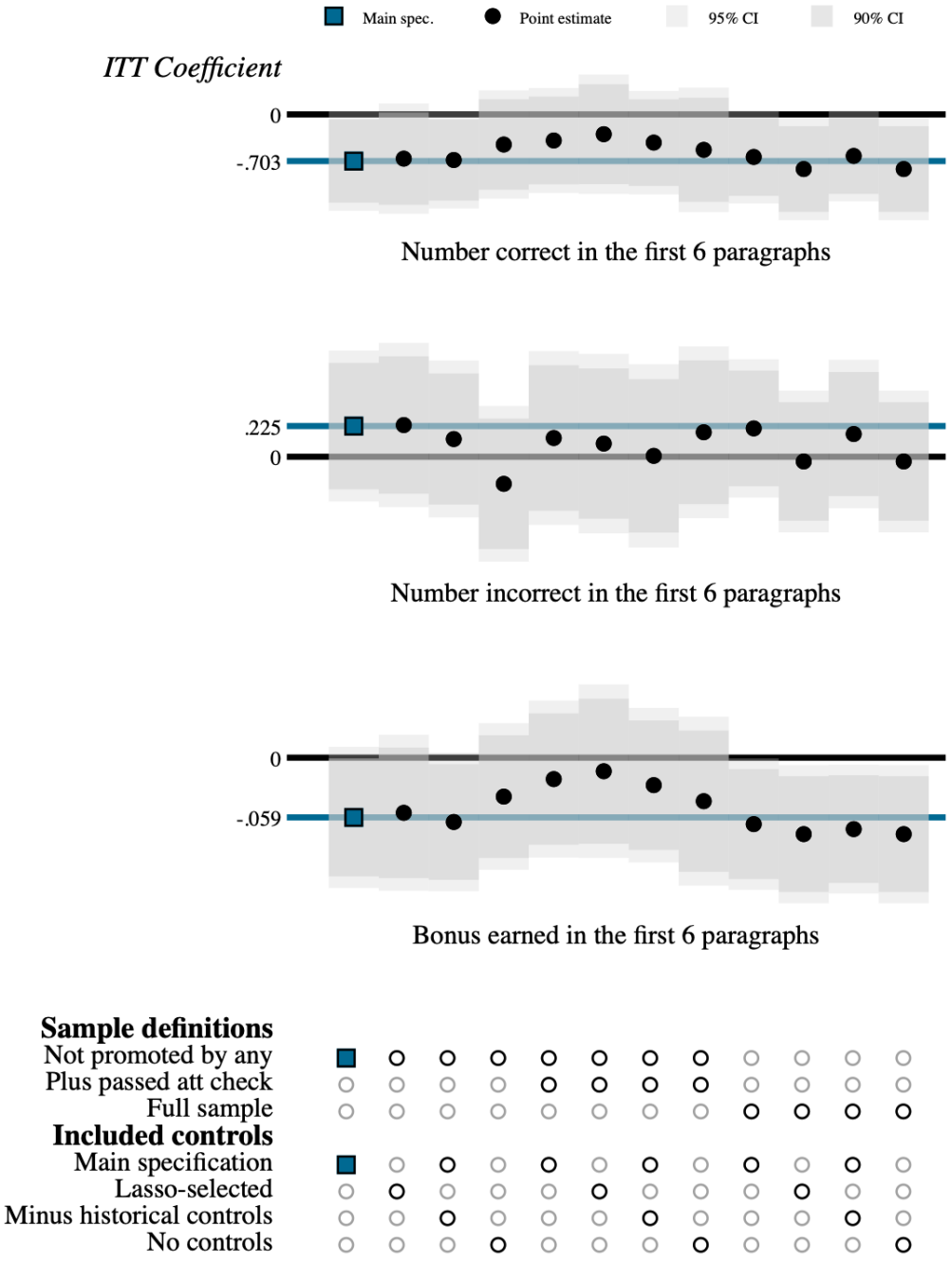
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on retention in the avatar arm of the promotion experiment. The outcomes and main specification are those in Table 2. Otherwise, the figure is analogous to Appendix Figure G16.

Figure G18: Alternative specifications: Effects of seeing avatars in the algorithm arm on effort



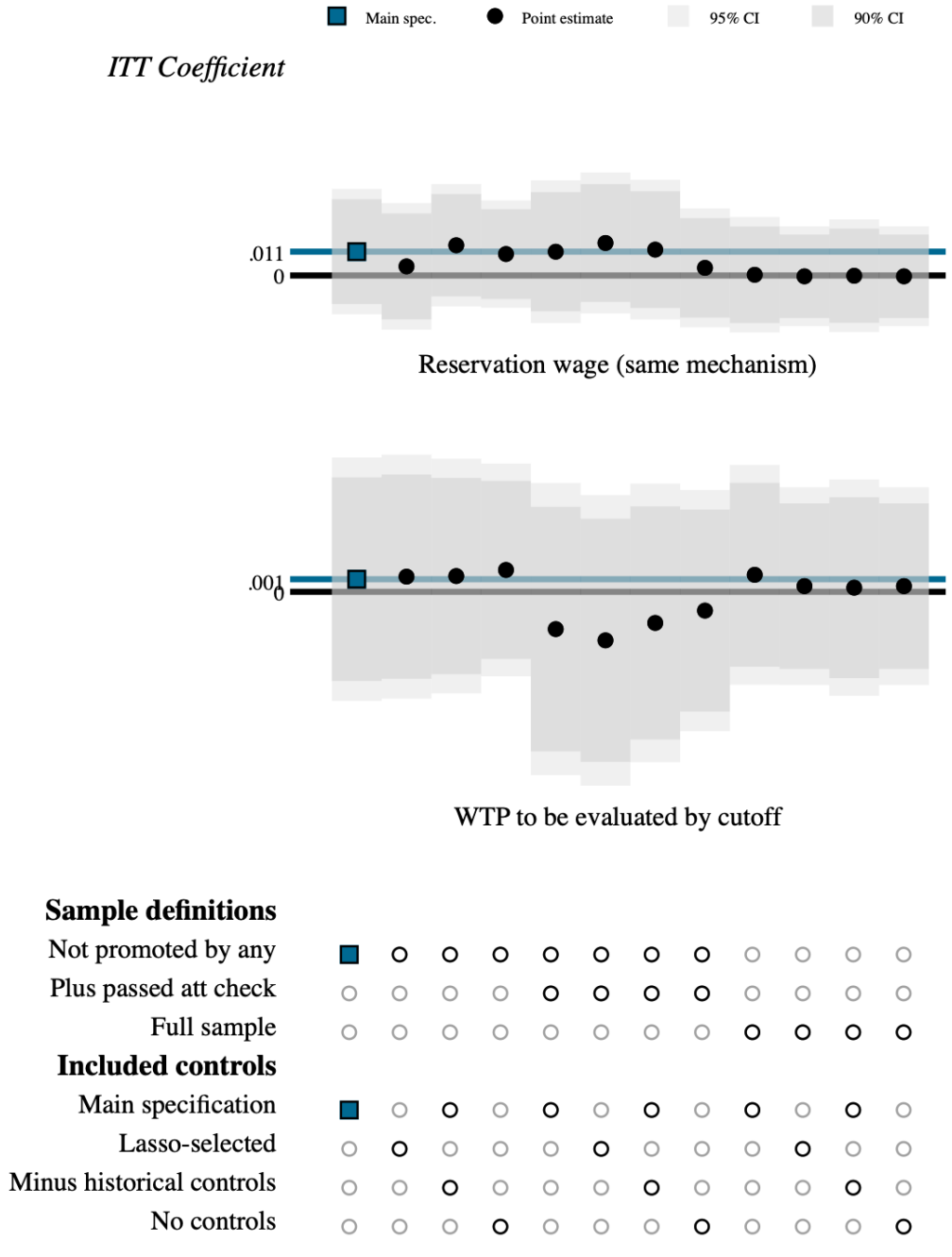
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on effort in the avatar arm of the promotion experiment. The outcomes and main specification are those in columns 1 and 2 of Table 3. Otherwise, the figure is analogous to Appendix Figure G16.

Figure G19: Alternative specifications: Effects of seeing avatars in the algorithm arm on performance



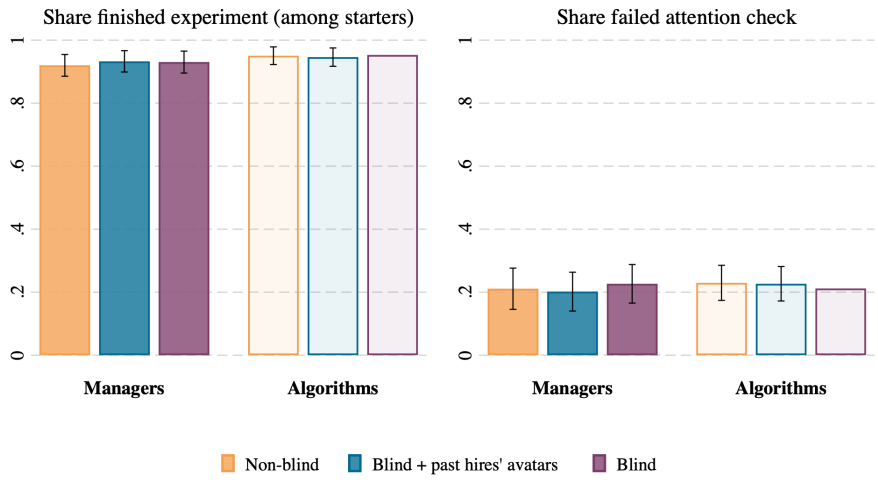
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on performance in the avatar arm of the promotion experiment. The outcomes and main specification are those in columns 3-5 of Table 3. Otherwise, the figure is analogous to Appendix Figure G16.

Figure G20: Alternative specifications: Effects of seeing avatars in the algorithm arm on future labor supply



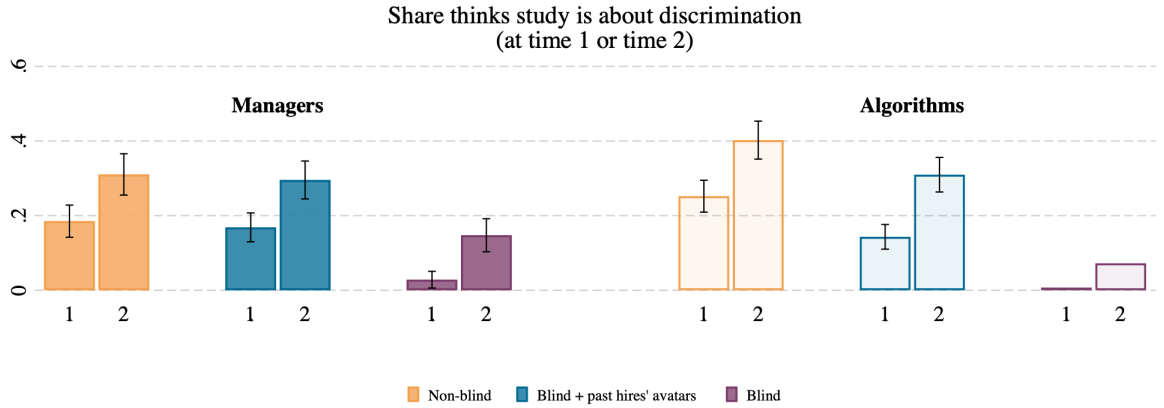
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on reservation wages in the avatar arm of the promotion experiment. The outcomes and main specification are those in Table 4. Otherwise, the figure is analogous to Appendix Figure G16.

Figure G21: Attrition and attention in the hiring experiment



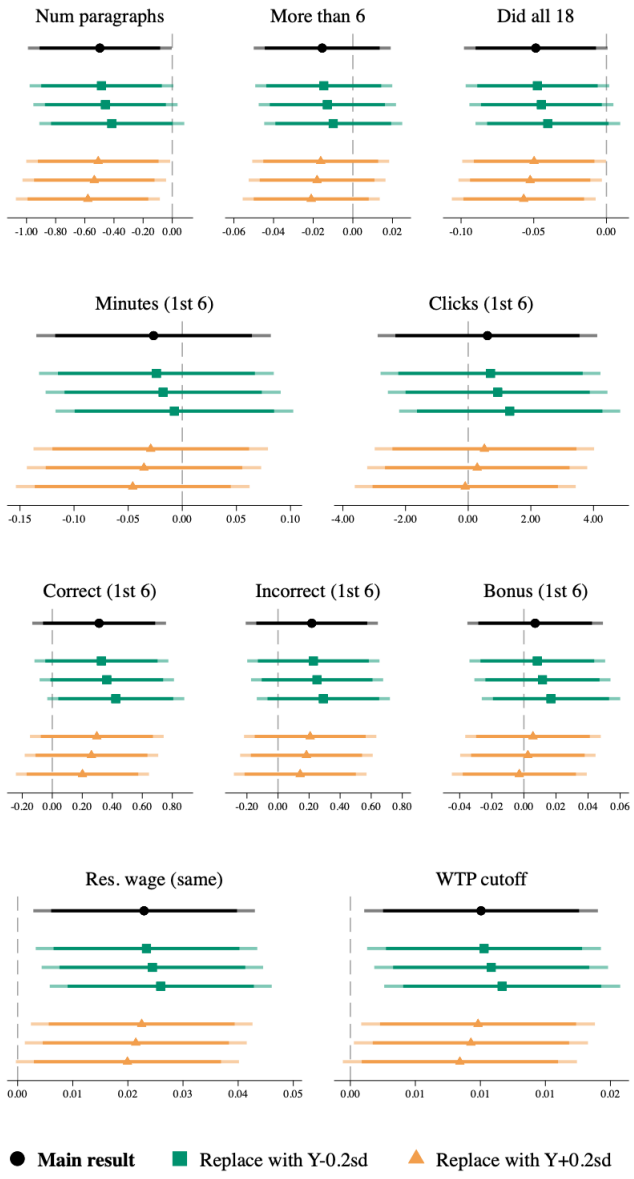
Note: This figure shows the fraction of workers in each arm of the hiring experiment who finished the experiment, among those who started it (on the left) and the share of the main sample that failed the attention check (on the right). The sample is restricted to workers who would not have been promoted 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 regressions that 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. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

Figure G22: Knowledge of study topic in the hiring experiment



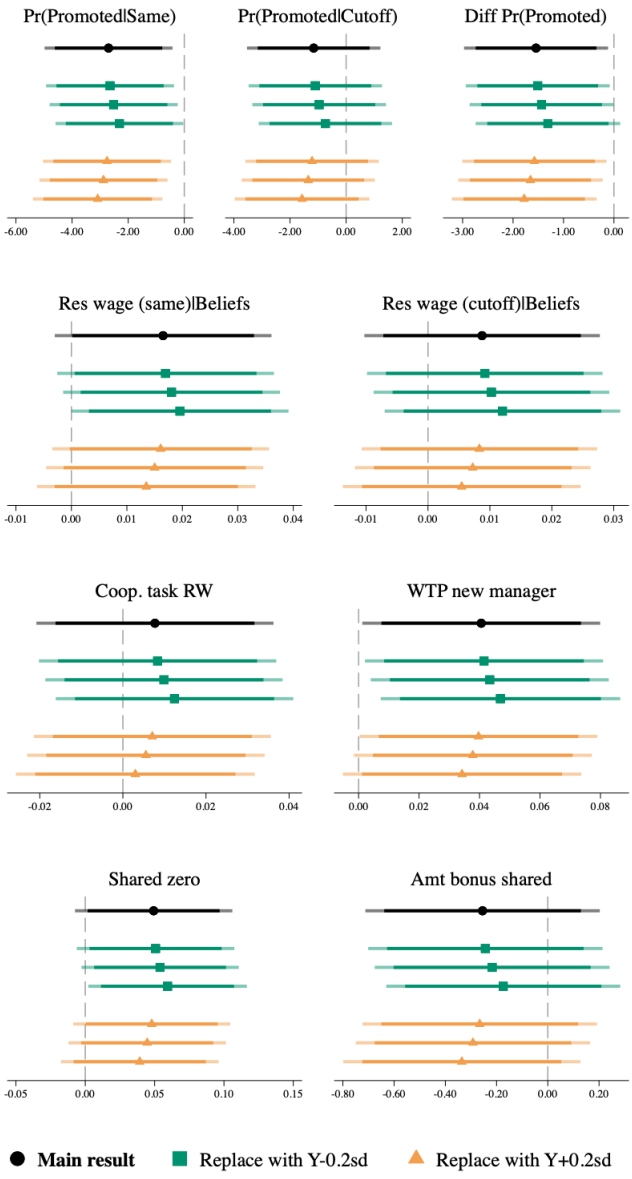
Note: This figure shows the fraction of workers in each arm of the hiring experiment who thought that the study was about discrimination the first time they were asked (after learning that they weren't hired and the reservation wage elicitations, before any survey questions) and the second (at the end 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 regressions that 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. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.

Figure G23: Accounting for possible demand effects in the manager arms of the promotion experiment: effort and future labor supply



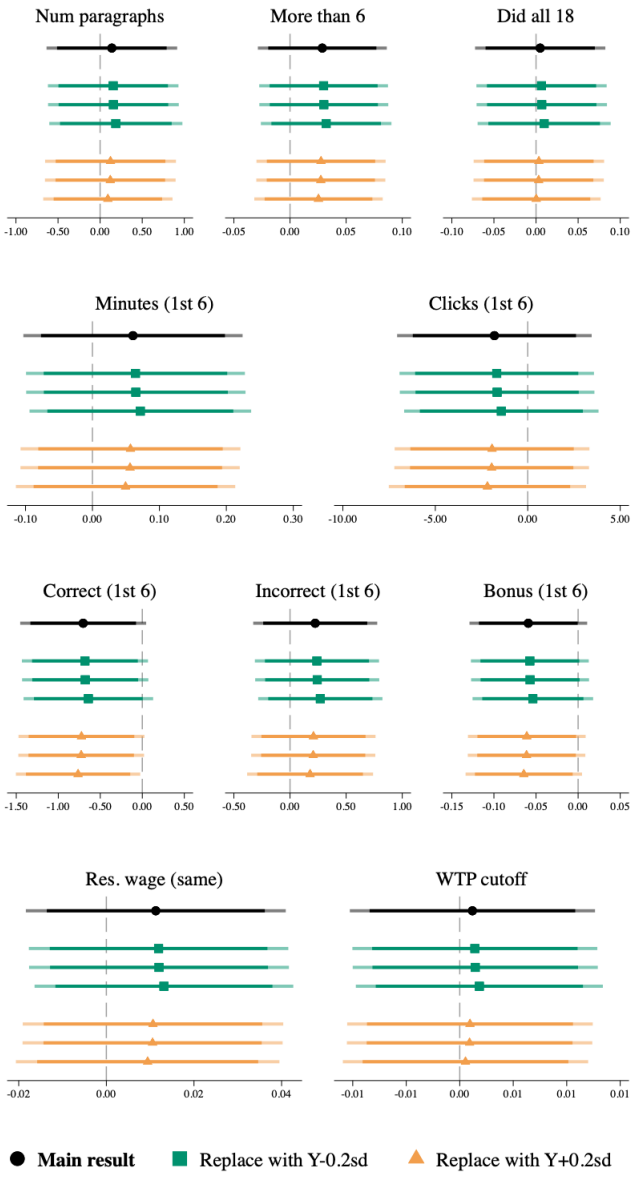
Note: This figure plots the treatment effects of being in the non-blind manager arm in the promotion experiment, relative to the demographic-blind manager arm, for the outcomes in Tables 2, 3, and 4, with adjustments to account for possible experimenter demand effects. These bounds replace the outcome 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 G24: Accounting for possible demand effects in the manager arms of the promotion experiment: secondary outcomes



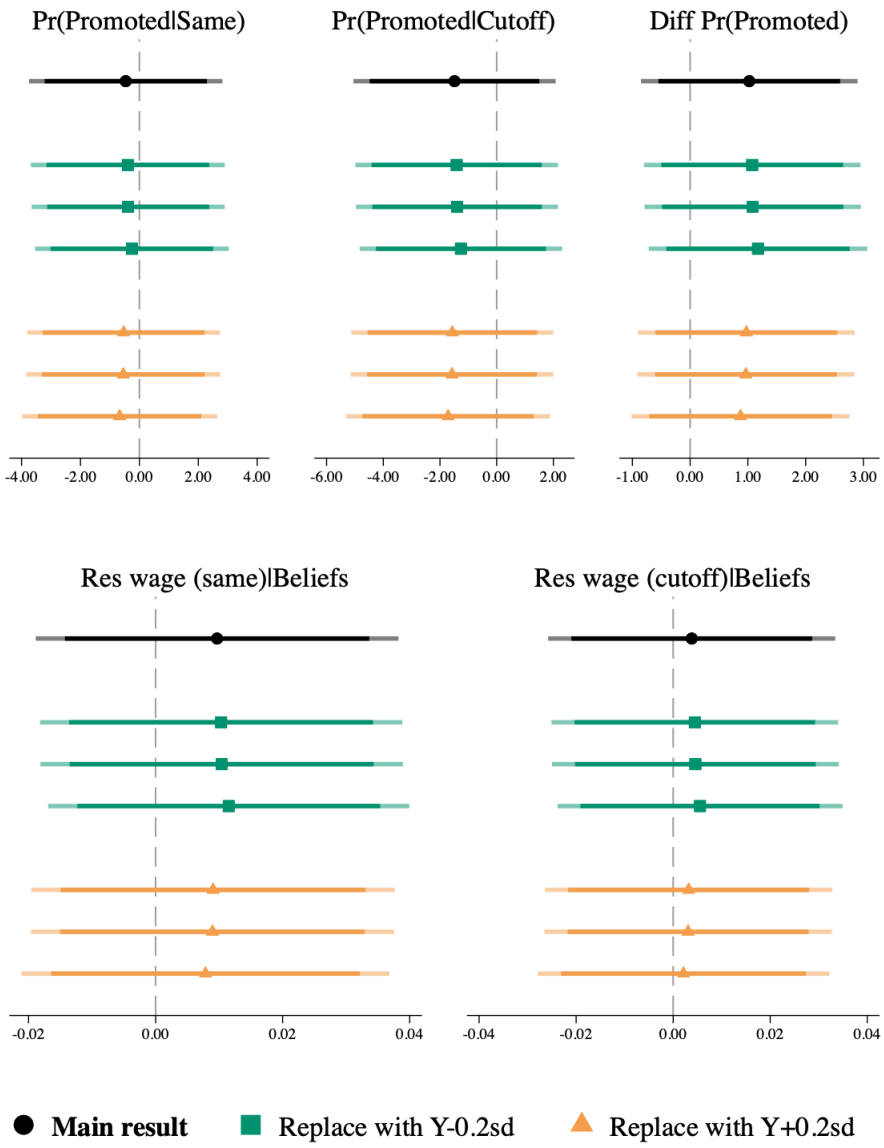
Note: This figure plots the treatment effects of being in the non-blind manager arm in the promotion experiment, relative to the demographic-blind manager arm, for the outcomes in Tables 6 and 5, and Figure 3, with adjustments to account for possible experimenter demand effects. The figure is otherwise analogous to Appendix Figure G23.

Figure G25: Accounting for possible demand effects in the algorithm arm of the promotion experiment: effort and future labor supply



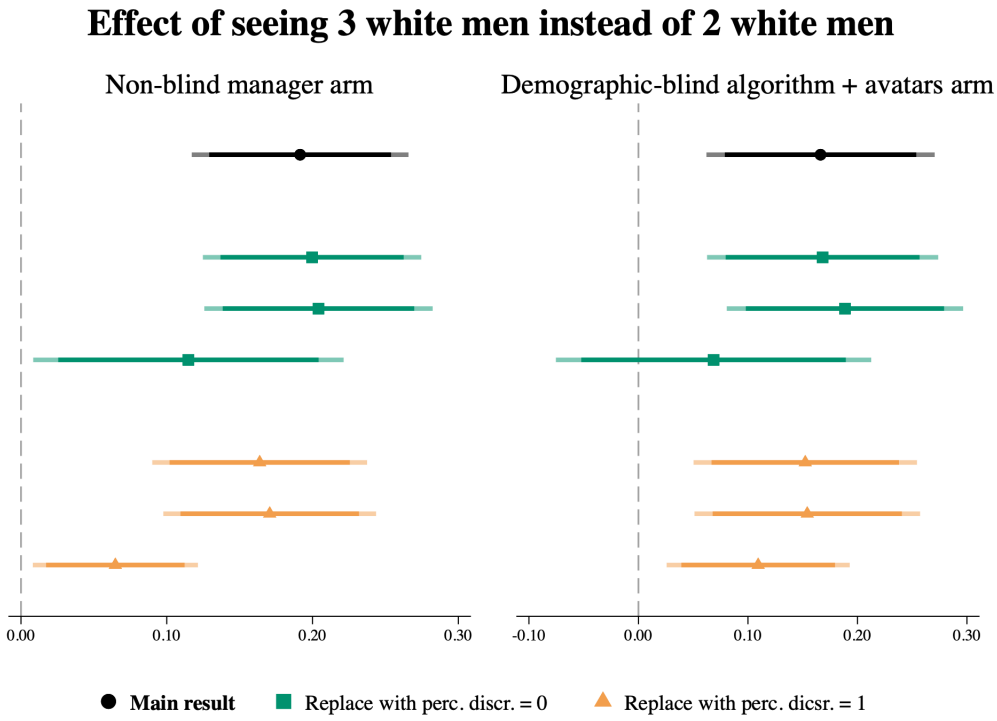
Note: This figure plots the treatment effects of learning the demographics of previously-promoted workers in the promotion experiment among workers evaluated by the demographic-blind algorithm with adjustments to account for possible experimenter demand effects. These bounds replace the outcome variable with its value plus or minus 0.2 standard deviations of that variable (for workers in the algorithm arm who do not learn the demographics of previously-promoted workers) 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 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. The three estimates for each replacement outcome variable use workers' ideas of the study topic at its first, second, and third elicitation, respectively. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure G26: Accounting for possible demand effects in the algorithm arm of the promotion experiment: secondary outcomes



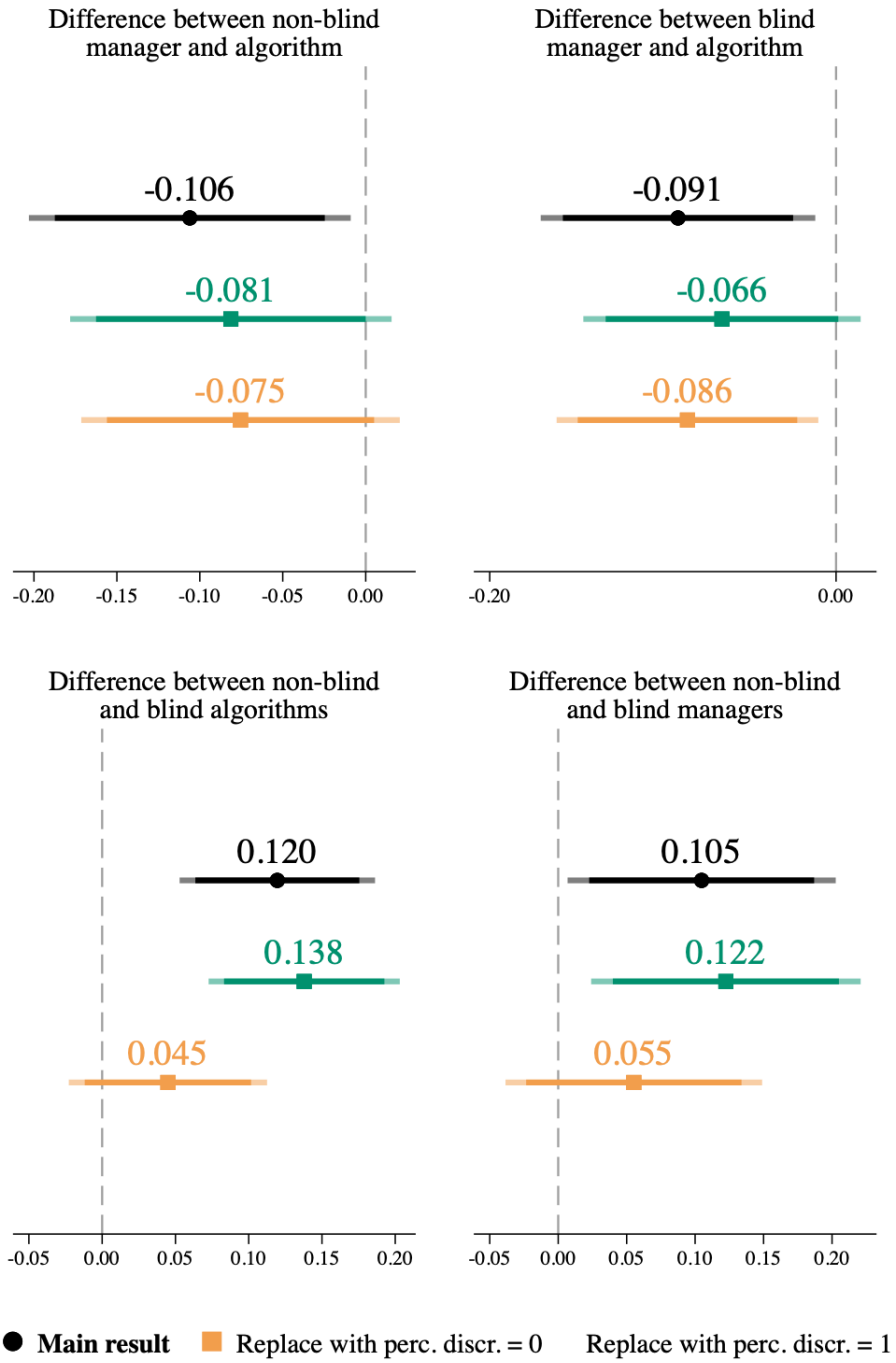
Note: This figure plots the treatment effects of learning the demographics of previously-promoted workers in the promotion experiment among workers evaluated by the demographic-blind algorithm with adjustments to account for possible experimenter demand effects, analogous to Appendix Figure G25 for the secondary outcomes in Figure 7.

Figure G27: Accounting for possible demand effects among workers who saw three versus two white men previously promoted



Note: This figure plots the effects of seeing that three white men were previously promoted from Figure 5, 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). 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 whether they saw that two or three white men were previously promoted. The specifications are otherwise identical to Figure 5. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Figure G28: Accounting for possible demand effects in the hiring experiment



Note: This figure plots the differences in perceived discrimination between treatment arms in Figure 4, 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 differences between treatment arms 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 Figure 4. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

Table G1: 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	-0.011	-0.002	-0.003
	(0.018)	(0.012)	(0.024)
N	1735	1450	1389
Control mean	0.837	0.953	0.239
Panel B: Algorithm arms of the promotion experiment			
See avatars	-0.027	0.033*	-0.044
	(0.028)	(0.018)	(0.040)
N	859	727	691
Control mean	0.860	0.932	0.275
Panel C: Non-blind manager arm of the promotion experiment			
See three white men	-0.036	0.012	0.015
	(0.026)	(0.014)	(0.037)
N	865	722	695
Mean, see two white men	0.851	0.954	0.223
Panel D: Algorithm+avatars arm of the promotion experiment			
See three white men	-0.040	-0.020	0.156**
	(0.044)	(0.024)	(0.062)
N	396	329	320
Mean, see two white men	0.844	0.977	0.213

Note: This table tests for differential attrition in the promotion 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 experiment. In the third column, 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 levels indicated by *, **, and ***, respectively.

Table G2: 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	0.030***	0.085***	0.101***
	(0.011)	(0.022)	(0.029)
N	697	692	1389
Control mean	0.011	0.038	0.399
Panel B: Algorithm arms of the promotion experiment			
See avatars	0.039**	0.042**	0.053
	(0.019)	(0.016)	(0.042)
N	353	338	691
Control mean	0.000	0.000	0.353
Panel C: Non-blind manager arm of the promotion experiment			
See three white men	0.065***	0.053	0.086*
	(0.023)	(0.047)	(0.046)
N	344	351	695
Mean, see two white men	0.005	0.132	0.449
Panel D: Algorithm+avatars arm of the promotion experiment			
See three white men	0.022	0.090**	-0.009
	(0.051)	(0.045)	(0.067)
N	161	159	320
Mean, see two white men	0.018	0.038	0.389

Note: This table tests for differential knowledge of the study topic in the promotion 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 levels indicated by *, **, and ***, respectively.

Table G3: 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	0.082	0.033	*0.006*
Failed att check 1 (screener)	0.087	0.062	*0.244*
Failed att check 2 (screener)	0.120	0.058	*0.005*
Failed att check (experiment)	0.361	0.184	*0.000*
Education:			
Less than high school	0.019	0.008	*0.217*
High school graduate	0.120	0.143	*0.432*
Some college but no degree	0.279	0.293	*0.699*
2 year college degree	0.106	0.099	*0.792*
4 year college degree	0.399	0.349	*0.212*
Professional or Masters degree	0.072	0.099	*0.257*
Doctorate	0.005	0.008	*0.623*
Proofreading performance, effort:			
Correct highlights, required paragraphs	12.471	15.806	*0.000*
Incorrect highlights, required paragraphs	4.346	3.017	*0.000*
Time on required paragraphs	4.872	4.992	*0.130*
N (blind manager arm)	208	484	
Panel B: Treatment effects (SEs)			
Number of paragraphs	0.234 (0.469)	-0.796 (0.281)	*0.050*
Proofread more than required	0.041 (0.034)	-0.037 (0.019)	*0.032*
Finished all paragraphs	0.011 (0.047)	-0.074 (0.028)	*0.104*
Time on required paragraphs	0.081 (0.102)	-0.068 (0.062)	*0.191*
Bonus in required paragraphs	0.038 (0.040)	-0.009 (0.024)	*0.306*
Beliefs about promotion, same mgr	-4.284 (1.980)	-1.917 (1.341)	*0.301*
Beliefs about promotion, cutoff	-4.701 (2.016)	0.304 (1.422)	*0.034*
Pr(same manager) - Pr(cutoff)	0.417 (1.251)	-2.221 (0.830)	*0.066*
Shared zero with manager	0.159 (0.051)	0.003 (0.033)	*0.008*
Amount shared with manager	-0.770 (0.436)	-0.033 (0.263)	*0.134*
N (both manager arms)	402	985	

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. Workers are dropped if they display multiple switching points or switch in the wrong direction (e.g. display downward-sloping labor supply) for any of the three reservation wage elicitations or the elicitation of willingness to pay to be able to choose one's own manager. 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. The specification is the same as Table 2. In panel B, standard errors are robust to heteroskedasticity (in parentheses).

H Coding the text data

As the first step for coding all three variables that came from open-ended responses, I read every response and coded them myself. These measures are not used in the paper, but correlate extremely highly with the final measures. To ensure that the process was entirely blind to the study purpose and treatment assignment, external coders generated equivalent measures and these are what are used in the paper. In the promotion experiment, the measures used in the paper were coded by two MIT undergraduate economics majors and their disagreements were resolved by a professional hired on Upwork (all three blind to study purpose and treatment assignment). The process with these coders was as follows:

1. Given my initial reading, I generated coding schemes (described in each subsection below) for each variable and provided them in an Excel spreadsheet with all of the open-ended responses, stripped of all identifiers and other variables.
2. The two undergraduate coders each received the same spreadsheet and separately coded each variable according to the coding scheme using a drop-down menu in Excel. Each variable was a different sheet in the spreadsheet. When they had questions, the questions and my response were shared with both of them, but they did not otherwise communicate about the exercise. They could attach up to three codes to each response answer.
3. Any observations with a discrepancy between the two coders (details on frequency follow) were reviewed by a PhD psychologist who professionally codes qualitative data hired on Upwork. Discrepancies included having the same codes but in a different order or having a different set of codes from each other. She could either indicate agreeing with one of the two original coders or she could re-code the response (which generally meant combining/re-ordering the original coders' tags).

In the hiring experiment, due to budget and time constraints, I only hired the professional coder on Upwork to code the data. Now, she was carrying out the same process as the undergraduate coders did for the data from the promotion experiment. Again, her measures correlate extremely highly with those that I coded myself. The codebooks for each variable were the same in both experiments and are described next.

In the promotion experiment, the external coders generated the indicators used in the paper for the main measure of perceived discrimination, whether workers complained about bias, and whether they knew the study purpose. In the hiring experiment, due to budget and time constraints, the external coder generated the indicators used in the paper for the main measure of perceived discrimination and whether workers complained about bias, but I generated the indicators for whether workers knew the study purpose myself.

H.1 The main measure of perceived discrimination

Reasons workers gave for their perception of what needed to be different about their profile to be assigned to the harder job could be classified into one of the following five categories, with the descriptions below provided to each coder. They ranked up to three given reasons (no observation had more than three reasons mentioned).

1. **More school**: They reference needing more school or a higher degree
2. **Higher quiz score**: They reference needing a higher score, higher quiz score, or "more stars"

3. **Demographics**: They say they would have been promoted if they had (a) different demographic(s) characteristic(s) or reference bias or discrimination

4. **Random**: They say it was random, or they don't know because they have the same qualifications as the others

5. **Other**

The undergraduate coders disagreed for 317 of 2,397 observations and these were resolved by the coder on Upwork. Of these, she agreed with coder 1 for 33 percent of observations, coder 2 for 58 percent of observations, and provided new codes for 8 percent of observations.

The main measure of perceived discrimination is an indicator for whether the final code was that the worker mentioned demographics as a reason they weren't promoted, regardless of its rank in their reasons. Of the workers who mentioned demographics at all, 58 percent mentioned it as their primary reason.

The control group means and treatment effects of evaluation by a non-blind manager on whether the *primary* reason was demographics, more school, higher quiz scores, or random are in Appendix Tables A5 and A15 (Panel A) and Appendix Tables A10 and A16 (Panels A, B, and C), respectively.

H.2 Secondary measure of perceived discrimination (complaints about bias)

Workers were asked near the end of the survey if they had any complaints about how they were assigned to jobs. If they said yes (N=368, 15 percent of workers), they were asked to describe their complaint. Workers' complaints could be classified into one of the following five categories, with the descriptions below provided to each coder. They ranked up to three reasons the worker had a complaint.

1. **Discrimination**: On the basis of any demographics

2. **Unfair**: Any unfairness other than discrimination on the base of demographics

3. **College degree**: College degree is irrelevant and shouldn't have been used to make promotion decisions, or a worker "was discriminated against because they didn't have a degree."

4. **Confusion**: They seem confused about the job or job-assignment mechanism, or don't remember taking the screening quiz

5. **Other**

The undergraduate coders disagreed for 128 of 368 observations and these were resolved by the coder on Upwork. Of these, she agreed with coder 1 for 35 percent of observations, coder 2 for 61 percent of observations, and provided new codes for 4 percent of observations. The complaints-based measure of perceived discrimination is an indicator for whether the final code was that the worker complained about discrimination, regardless of its rank. Of those who complained about discrimination, it was the primary complaint for 93 percent (only 33 workers gave more than one reason for their complaint). Of those who had a complaint, 30 percent referenced discrimination in their description (23 percent referenced other types of unfairness, 18 percent referenced college degrees, and 10 percent were confused; 29 percent had some other complaint).

The treatment effects on whether workers had a complaint about bias are in Appendix Tables A5 and A15 (Panel C).

H.3 Study topic

Workers were asked about their beliefs about the study topic at two times in the survey. In the proofreading experiment, half of the sample was randomly chosen to be asked for the first time between the proofreading section and the reservation wage and willingness to pay elicitations, the other half was asked after the reservation wage elicitations, and all workers were asked a second time at the very end of the survey. In the hiring experiment, all were asked after the reservation wage elicitations and at the very end of the survey. Their beliefs about the study topic could be broadly categorized into six categories, with the following descriptions given to each coder. The coders ranked up to three codes per response (no response referenced more than three topics).

1. **Proofreading**: They think the study is about proofreading, including gender, race, or education differences in proofreading ability
2. **Motivation/persistence**: They think the study is about persevering through a long and difficult or tedious job
3. **Perceptions of job assignment**: They think the study is about perceptions of job assignment but don't mention discrimination or bias specifically. Can include fairness.
4. **Discrimination/bias**: They think the study is about (perceptions of) discrimination or bias in job assignment
5. **I don't know**: They don't know
6. **Other**

Study topic 1: The undergraduate coders disagreed for 933 of 2,397 observations and these were resolved by the coder on Upwork. Of these, she agreed with coder 1 for 38 percent of observations, coder 2 for 52 percent of observations, and provided new codes for 10 percent of observations. Study topic 2: The undergraduate coders disagreed for 885 of 2,397 observations and these were resolved by the coder on Upwork. Of these, she agreed with coder 1 for 55 percent of observations, coder 2 for 33 percent of observations, and provided new codes for 12 percent of observations.

Appendix Table G2 shows treatment effects on whether workers mentioned discrimination/bias. Of those who mentioned discrimination/bias when asked what they thought the study topic was the first time (N=102), 67 percent mentioned discrimination primarily. Of those who mentioned discrimination/bias at the end of the survey (N=1013), 79 percent mentioned discrimination primarily.

In the promotion experiment, the primary study topic participants hypothesized the first time and second time they were asked followed this distribution:

	1st time asked	2nd time asked
Proofreading	28 percent	11 percent
Motivation	16 percent	5 percent
Perceptions of assignment	12 percent	13 percent
Discrimination/bias	3 percent	33 percent
I don't know	31 percent	30 percent
Other	10 percent	8 percent

The share of workers who thought the study was about discrimination in the hiring experiment is plotted in Appendix Figure G22.

I Effects on future labor supply in the hiring experiment

Unlike in the promotion experiment, both perceived manager and algorithmic discrimination increase beliefs about the likelihood of future discrimination (thus reducing beliefs about the likelihood of future job offers) when a worker will be evaluated by the same procedure in the future. There are two possible explanations for this difference between experiments. The difference may be due to differences in the timing of the reservation wages and belief elicitations relative to when workers learned about how they were evaluated, or due to the size of the "first stage," both of which differ between the two experiments and make it more likely for there to be a larger effect on anticipated future discrimination in the hiring experiment. In Panel A of Table I1, I pool the two manager arms with positive perceived discrimination and compare to the demographic-blind manager arm where workers do not see previous hires' avatars, race, and gender; in Panel B, I do the same in the algorithm arms. I pool the arms to improve power (as pre-registered), but results are similar in each arm separately (Appendix Figure I1). Column 1 shows the effect of treatment on perceived discrimination in the pooled treatments (36pp and 53pp when evaluated by a manager or algorithm, respectively, se=2pp for both). Being in the manager arms with perceived discrimination lowers beliefs about the probability of being hired by 2.5pp (5 percent, se=1.2pp) in the manager arms or 7.8pp (15 percent, se=1.1pp) in the algorithm arms (Table I1, column 2). 90 percent and 51 percent of these effects, respectively, remain after subtracting workers' beliefs about the likelihood of promotion under the cutoff rule, and is therefore due to increased anticipated discrimination (column 3).

I also measure reservation wages for future work in the hiring experiment. Effects on these reservation wages are purely indicative of psychological costs from anticipated discrimination: workers' choices depend only on one wage, not an expected wage, so movement of their switching point along the multiple price list cannot be caused by changes in beliefs about the likelihood of being hired.

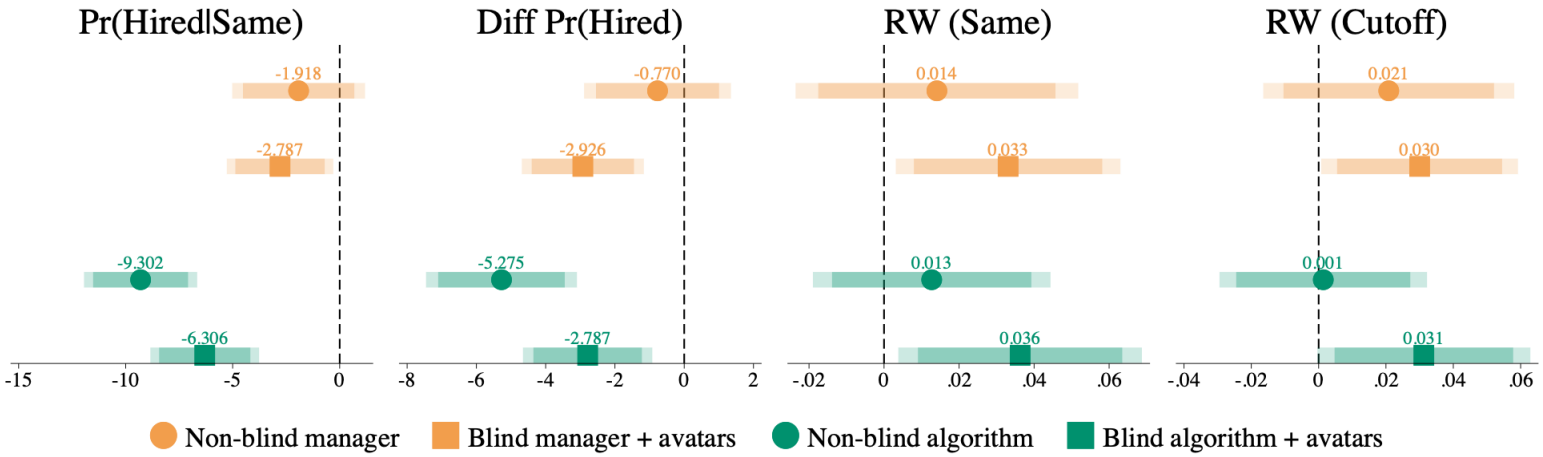
Workers do have higher reservation wages when they perceive (and anticipate) discrimination, *even though they will not interact with any manager or employer in the future if they are* ***not*** *hired.* That is, they only want to be offered the job at higher wages, implying a direct utility cost from anticipated discrimination or future interactions, just like in the promotion experiment. Unlike in the promotion experiment, this seems to be largely due to avoidance of the employer who hires biased managers or uses biased algorithms, because reservation wages increase similarly when workers will be evaluated by the same procedure as in the experiment (Table I1, column 4) as when they will be evaluated by a cutoff rule in baseline performance (column 5). As with the promotion experiment, these results imply that workers will sort in or out of jobs based on anticipated discrimination and require higher wages in situations where they may face discrimination. Consistently, I find suggestive evidence that the treatments worsen workers' psychological well-being.

Table I1: Effects on future labor supply and beliefs in the hiring experiment

	Perceived Discrimination	Pr(Hired) (Same mech)	Diff Pr(Hired) (Same-Cutoff))	RW (Same)	RW (Cutoff)
	(1)	(2)	(3)	(4)	(5)
Panel A: Manager sample, hiring experiment					
Non-blind manager	0.361***	-2.516**	-2.252***	0.027*	0.027**
or blind mgr + avatars	(0.020)	(1.166)	(0.827)	(0.014)	(0.014)
N	1246	1246	1246	1231	1228
Control mean	0.007	51.572	-2.418	0.388	0.383
Panel B: Algorithm sample, hiring experiment					
Non-blind algorithm	0.529***	-7.753***	-3.988***	0.025*	0.017
or blind alg + avatars	(0.018)	(1.119)	(0.869)	(0.014)	(0.014)
N	1278	1278	1278	1254	1254
Control mean	0.002	51.545	-2.762	0.414	0.415

This table reports treatment effects on perceived discrimination, future labor supply, and beliefs about the likelihood of being hired in future.The outcome in the first column is the main measure of perceived discrimination. The second and third outcomes concern beliefs about being hired in the future: the probability that workers think they will be hired in the future if they are evaluated by the same procedure as in the experiment (column 2) and the same probability minus the probability they assign if they will be evaluated by a cutoff rule in baseline quiz scores instead (column 3). In columns 4 and 5, the outcome is their reservation wage per paragraph to be evaluated again and potentially hired if they are evaluated by the same procedure as in the experiment (column 4) or the cutoff rule (column 5). In panel A, the sample is those evaluated by a manager and the presented coefficient is on an indicator for being in *either* the non-blind manager arm or the demographic-blind manager arm where workers see previous hires' avatars, race, and gender, rather than the demographic-blind manager arm that does not see previous hires' avatars, race and gender. Panel B is the analogous specification in the three algorithm arms. Appendix Figure I1 shows results that do not pool the arms in this way. Both samples restrict to workers who would not have been hired under any procedure. The effect on perceived discrimination is different from Figure 4 because the sample does *not* restrict to workers who saw three white men previously hired. 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 levels indicated by *, **, and ***, respectively.

Figure I1: Effects on beliefs, reservation wages, and affect in the hiring experiment, separately by treatment arm



Note: This figure plots the estimated treatment effect of being in each arm with positive rates of perceived discrimination in the hiring experiment. The outcomes, specifications, and samples are the same as in Table I1—the only difference is that the effect of the two manager arms with positive perceived discrimination are estimated separately relative to the demographic-blind manager arm that does not see previous hires' avatars, race, and gender, and analogously in the algorithm arms. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.