Perceived discrimination at work

Hannah Ruebeck*

MIT

December 1, 2023

You can find the most current version here.

Abstract

Minority-group workers in the US understand that discrimination is widespread, and when they experience rejection at work, observe incomplete information about the cause. Whether accurate or not, perceived discrimination may generate distrust, retaliation, or anticipated further discrimination. It is therefore important to assess how workers form beliefs about experiencing discrimination and how those beliefs affect job performance and labor supply. In an online experiment (N≈5000), I randomly assign workers to be evaluated by promotion procedures with varied potential to discriminate and the information workers have about past promotions. When rejected job candidates learn that previously-promoted workers are primarily white men, many infer bias even when they know that hiring managers (or algorithms) did not know race and gender. These perceptions have important consequences: learning that a manager knew workers' race and gender and previously promoted mostly white men increases the share of workers who perceive discrimination from 1 to 31 percent, lowers retention by 3-6 percent, and increases reservation wages by 9 percent. My findings suggest that employers could shrink retention gaps and boost performance by reducing perceived discrimination, but that objectively neutral procedures are unlikely to alone eradicate perceptions of 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, the participants in MIT's labor and behavioral lunches, and seminar participants at Wellesley College for their thoughtful advice and helpful comments. 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's resesarch fund, 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 with 28 percent funded by ACF/HHS and $65,178 and 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 preregistration 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, experience outcomes consistent with discrimination, and often infer they have been discriminated against. In the last four decades, social scientists have documented widespread workplace discrimination against women and racial minorities (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 it is hard to measure and 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, morale, or anticipated further discrimination (Goldsmith et al., 2004).

Workers reasoning about whether they have experienced discrimination face a complex problem. Consider the care with which social scientists design and implement the audit studies that document discrimination at scale in the workplace (e.g. Kline et al., 2021). Individuals, on the other hand, receive few and 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.

This paper describes two large, pre-registered experiments (N$\approx$5000) testing how workers form beliefs about experiencing discrimination and how those beliefs affect retention, performance, and future labor supply. To do this, I set up a dynamic labor market with multiple rounds of worker evaluation and work opportunities on Prolific, an online platform where workers take surveys for pay. Workers are randomly assigned to be evaluated under procedures that differ in their 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 focus on gender and racial discrimination in a white- and male-dominated domain (science), and over-sample women and racial minority men. By observing whether workers would have been promoted under any of the procedures and restricting the sample to those who would never be promoted, I eliminate

[1] An exception is Gagnon et al. (2022), discussed in more detail 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 the 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 A34).

potentially confounding effects of actual discrimination.

The experiments reproduce a common scenario. Someone working in a field in which her demographic group is a minority applies for a job. She knows the hiring manager is in the majority group (e.g., from an interview) and that he previously hired other members of the majority group for this job (e.g., from the company's website). When she is rejected, she tries to figure out why. How often does this worker perceive discrimination?

I introduce a new measure of perceived discrimination that circumvents concerns with other methods. Prior surveys prime workers by asking directly about discrimination or differential treatment (e.g. Kessler et al., 1999; Goldsmith et al., 2004; Mukerjee, 2014; Gagnon et al., 2022). 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 selected. I define perceived discrimination when a worker's open-ended response mentions demographic information. My measure correlates strongly with more explicit measures asked at the end of the survey.

The experiments answer two research questions. First, I identify the implications of hiring practices that aim to mitigate the influence of human biases but have ambiguous implications for the prevalence of *perceived* discrimination. What if an algorithm ranked candidates and the highest-scoring ones were hired, or decision-makers did not know candidates' race and gender? I answer these questions in a setting where 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 randomly vary whether previously-hired workers are mostly or entirely white men to understand how minority-group representation affects these inferences. These 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).[3]

Second, my experiments investigate how perceived discrimination affects retention, performance, and future labor supply using the variation induced by these manipulations. I refer to workers' persistence in the job task after the treatment as *retention* to distinguish it from their

[3]Algorithms can embed human biases but are easier to regulate and audit than human decision-making, perhaps leading to less discrimination (e.g. Kleinberg et al., 2020) and thus less perceived discrimination, but are also 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 Kang et al. (2016) show that racial minorities are less likely to "whiten" their resumes when job ads emphasize diversity but that employers who do so are actually no less likely to discriminate against "unwhitened" resumes. Blinding decision-makers to demographic information can effectively improve representation of minority groups when they are under-represented (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, though the majority of practitioners are familiar with it (Fath and Zhu, 2021). Since it is not very common, employees learning about this practice may not believe the information.

labor supply to future job opportunities offered in the second phase of the experiment. I also consider whether these effects depend on whether managers or algorithms make decisions. Algorithmic bias is generated by statistical processes, not explicit prejudice, which may change how perceived discrimination affects behavior (Bigman et al., 2023).

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 promotion procedures vary in their potential to discriminate: (i) a manager who sees prior performance and education, (ii) a manager who *also* sees avatars, race, and gender, and (iii) an algorithm that predicts performance using prior performance and education. The randomly-assigned procedure's decision is implemented 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 that the only differences between treatment arms are workers' perceptions of why they were not promoted.[4]

What workers learn about how they were evaluated generates differences in perceived discrimination. Before commencing the easier job, non-promoted workers see how they were evaluated, including the decision-making inputs and profiles of the three workers who were promoted under that procedure in the past—primarily white men due to their over-representation in the earlier sample. 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. Workers evaluated by a non-blind manager see that the previously-promoted workers are primarily white men but workers evaluated by a demographic-blind manager do not, since they all see the same profiles their manager saw. I use this random variation to test the effect of increasing minority-group representation among previously-promoted workers. To generate variation in perceived *algorithmic* discrimination, workers evaluated by the (demographic-blind) algorithm are randomly assigned to see previously-promoted workers' profiles with or without avatars, race, and gender.

The second experiment (N≈2,700 new workers) disentangles the effects of using an algorithm versus a manager and the effect of blinding decision-makers to race and gender when workers see that primarily white men were previously hired. Unlike the first experiment, it follows a factorial design to separate these components.[5] I randomly assign workers to be evaluated by a manager or an algorithm and cross-randomize whether the decision-maker uses information about race and gender. Among workers evaluated by a demographic-blind decision-maker, I further cross-randomize whether they learn previous hires' race and gender. Again, primarily white men were

[4]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.

[5]See Figures 3 and 5 for an illustration of the differences.

previously hired for the job. The primary outcome is whether workers perceive discrimination. Now, 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.

I find that changes in hiring procedures that mitigate the direct effects of human biases—using algorithms or blinding managers to race and gender—do little to mitigate common *perceptions* of discrimination when past decisions "seem like bias," but that increasing minority-group representation effectively does so—an example of the representativeness heuristic affecting belief formation (Kahneman and Tversky, 1972; Mullainathan, 2002; Bordalo et al., 2021). Specifically, when workers see that only white men were previously hired and know their manager knew their race and gender, 52 percent perceive discrimination. Blinding managers to race and gender only reduces perceived discrimination by 11pp (20 percent, p-value=0.04) when white men are so overrepresented. Regardless of whether decision-makers can use demographic information or not, using an algorithm *increases* perceived discrimination by 9-11pp (10 percent, p<0.03). Using a demographic-blind algorithm reduces perceived discrimination relative to a non-blind manager by about 30 percent in the first experiment but does not reduce perceived discrimination in the second. These results suggest that transitions to algorithmic decision-making or efforts to de-bias managers can, at best, only moderately reduce perceived discrimination if company hires and leadership are not diverse. Seeing one woman or racial minority man among the three previously-selected workers is more effective: this increased minority-group representation reduces perceived manager and algorithmic discrimination by 45-56 percent (p<0.01).

Common, hard-to-change perceptions of discrimination are particularly concerning because I find that these perceptions affect retention and performance, which would exacerbate gaps caused by discrimination itself. In the first experiment, 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.[6] There are subsequently large effects on worker behavior: there are mixed effects on performance in the tasks required to receive the basic payment for the job but a substantial difference in labor supply for subsequent tasks. Being in the non-blind manager arm reduces the number of tasks completed by 3 percent (p=0.05). These effects are driven by women, for whom evaluation by a non-blind manager increases the probability of quitting as soon as they are able by 61 percent and lowers the chance of finishing the job by 7.5 percent (p=0.05, 0.04). Seeing that a demographic-blind algorithm previously promoted mostly white men—which increases perceived discrimination by 17pp—reduces performance by 5 percent (p=0.07) but does not affect retention.

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

[6]This is lower than above due to lower rates of previously-selected workers being all white men and a lower fraction of the sample being women, who are more likely to perceive discrimination in the experiment.

future interactions rather than lower expected wages due to future discrimination. Specifically, workers in the non-blind manager arm have 9 percent (p=0.01) 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 also 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, i.e. anticipated disutility from future discrimination or interaction with a biased manager. About a third of the 9 percent reduction is due to lower beliefs about the likelihood of promotion in the future. In other words, being in the non-blind manager arm increases reservation wages by 6 percent due to higher disutility from work, holding fixed expected wages. Consistently, perceived algorithmic discrimination has similar effects when workers anticipate future discrimination.

In sum, this paper shows that rectifying the disparate impacts of discrimination will require addressing worker perceptions, and this is due to both material and psychological costs. Furthermore, employers could shrink retention gaps and boost performance by reducing perceptions of 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; Kline et al., 2021; Bohren et al., 2019; Small and Pager, 2020; Bohren et al., 2022). I show that perceived discrimination affects minority-group worker behavior and can therefore exacerbate gaps caused by discrimination itself. In Bohren et al.'s framework (2022), perceived discrimination therefore contributes to systemic discrimination, affecting opportunities even if future evaluation is *prima facie* race- and gender-neutral. Prior work, therefore, understates the overall implications of discrimination in the labor market by neglecting the channel of perceived discrimination.

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 mechanisms play a substantial role. Thus, I build on prior 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, interview performance, and revelation of sensitive information (Alston, 2019; Charness et al., 2020; Lepage et al., 2022; Ridley, 2022; Aksoy et al., 2023; Angeli et al., 2023; Avery et al., 2023).[7] Unlike my paper, these do not identify whether effects are driven by changes in perceived returns or psychological mechanisms. Recent papers also show 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 varying potentially discriminatory interactions and show 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 dimension of behavior I call retention) relative to the effects of wage inequality alone. Rather than directly informing workers that discrimination occurred, I aim to mimic a real-world situation in which workers learn about firm hiring procedures and their results, and possibly infer discrimination. I provide evidence on what types of policies might reduce perceived discrimination and use a novel measure to document the prevalence of perceived discrimination. I also show that this 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]

[7] Other papers show that making stigmatized or stereotyped identities salient or advertising company diversity affects job application rates and task performance, but channels other than anticipated discrimination could explain these results (e.g. Hoff and Pandey, 2006; Kang et al., 2016; Flory et al., 2021; Burn et al., 2023; Del Carpio and Fujiwara, 2023). Pager and Pedulla (2015) show that Black workers cast a wider net in job search, and that this is positively associated with reporting experiences of discrimination at work in the past two years; Goldsmith et al. (2004) and Mukerjee (2014) show that controlling for past perceived discrimination can explain racial differences in labor supply and job satisfaction, respectively.

[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).

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 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.

2 Research design

The key challenge in studying the effects of perceived discrimination is that those who perceive discrimination will always be more likely to have experienced discrimination and its direct effects. These workers are also more likely to be from oft-discriminated groups, who are—observably and non-observably—different from other groups in many ways. With careful design considerations, 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—thus circumventing new challenges introduced in an experimental context.

The remainder of this section describes how I address these key 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, also 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 who are most 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) (University of Minnesota, 2021).

The experimental sample includes people representative of large swaths of the US population and almost all (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 per year, and the rest earn above $100,000 per year. 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, and 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. I restrict the ACS sample to individuals ages 18-71, the minimum and 99th percentile of age in the experimental sample, respectively, in column 3 and further restrict to employed women and racial

[10] 3,240 workers were recruited to the baseline survey for the first experiment. 120 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). Of the 3,120 remaining workers, 2,886 are not promoted by their randomly-assigned evaluation procedure and thus offered the chance to participate in the experiment, which 80 percent of them complete. Of those, 90 percent would not have been promoted under any procedure regardless of their random assignment, and thus make up the final analysis sample. 3,960 workers were recruited to the baseline survey for the second experiment. 360 are dropped from the experimental sample because the list of possible managers to evaluate them was exhausted (again, randomization occurred within these groups of 360). Of these, 3,510 were not hired under their randomly-assigned evaluation procedure and thus offered the chance to participate in the experiment, which 76 percent of them complete. Of those, 94 percent would not have been hired under any procedure and make up the final analysis sample.

minority men in column 4. The experimental sample is more likely to have a college degree, but also more likely to have household income in the $20,000-$60,000 range and less likely to earn above $160,000 than either of these representative samples.

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. 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.

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-level articles in the *Science Journal for Kids*. 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. Each paragraph has around 100 words and four inserted errors. Most workers use almost the entire allotted minute to proofread each paragraph. Once they've proofread the first six, workers choose after each paragraph whether to continue proofreading or go to the end of the survey.[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. In particular, they earn the bonus for a given paragraph if two times the number of correct highlights minus the number of incorrect highlights is greater than or equal to the median of the same statistic for that paragraph in a pilot. Workers are not told this precise rule, however—they are told 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.

This scientific proofreading job was chosen for several reasons. First, women and racial minority men are under-represented in most scientific fields—the context I was trying to create in my experiment. Second, of the effortful tasks that can be embedded in a survey experiment, it

[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.

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 vaguely 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 threshold-setting for payment. Empirically, however, workers anticipated the use of objective performance metrics when asked at the end of the survey how they thought "a good job" would be determined.

What did workers know about the jobs? Appendix Figure A1 contains the exact information provided to workers about the jobs, which they see at the beginning of the experimental survey before learning they were not promoted. 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 that is "well-proofread."[12]

2.2.2 Design overview and background

Figure 1 illustrates the overall experimental design. The key feature is that 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—use information on workers' prior performance and education to decide who to promote. Before recruiting the experimental sample, the same managers and algorithm evaluate groups of workers in which white men were over-represented (a timeline is in Figure 2). The key to inducing variation in *perceptions* of discrimination in the experimental sample is that workers learn about the procedure under which they were not promoted and who was previously promoted, with additional random variation in what workers know about these workers.

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] Workers are first told the precise wages in each job. Then, to ensure that workers prefer promotion to non-promotion, they are told that about half the paragraphs are proofread or summarized well in both jobs, i.e., that workers may have to exert more effort but would indeed earn more in the harder job. At the end of the study, 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 proxy for 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.[14] 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, thus seeing profiles of the same type as when they evaluated the historical sample. Each manager now 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). It was trained on a sample of 500 workers recruited from MTurk using Cloud Research's pre-approved participant pool, whose baseline quiz scores and performance in the harder job were collected in one survey. Like the experimental sample, women and racial minority men are over-represented in the training sample. Appendix C provides more detail about the algorithm.

2.2.3 Experimental design

This section describes the promotion experiment design, using the vocabulary defined and described in the previous section (i.e. *algorithm*, *demographic-blind manager*, *non-blind manager*, and *previously-promoted workers,* which refers to the three workers that each manager promotes in the historical sample). Figure 3 illustrates the treatment arms.

Baseline survey. Workers are recruited on Prolific for an initial survey, during which they take three timed quizzes that are predictive of their ability in the proofreading jobs and report how well they thought they did on each.[15] Workers know their performance on the quizzes could

to feel discriminated against and to minimize noise in the experiment caused by variation in manager characteristics (only 54 participated in the study). One manager self-identified as non-binary which workers saw rather Prolific's categorization of them as male. All managers were also required to have a job outside Prolific.

[14] Workers create an avatar that "looks like them" in the baseline survey, and are not told what it will be used for since this would have affected their choices (e.g. Alston, 2019; Charness et al., 2020; Ridley, 2022; Aksoy et al., 2023). There are more than 7,500 combinations of avatar characteristics (skin tone, hair color, hair style, glasses, facial hair, and gender—communicated via the color and style of the avatar's shirt).

[15] A spelling quiz of 10 common scientific words, 15 basic grammar questions, and 11 basic science questions. The

affect their future work opportunities and are paid a small bonus for each correct answer. Workers then answer demographic questions, build an avatar, and report a detailed work history, including past experiences of discrimination at work, and report their beliefs about what fraction of survey participants in eight race $\times$ gender groups will report previously experiencing discrimination.

Randomization and worker evaluation. Workers are randomly assigned to groups of 120 with average baseline quiz scores within one quintile of each other (e.g., all workers in a group are in the first or second quintile, or the second or third quintile, etc.). Each group is evaluated by two managers (randomly assigned), 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. Therefore, workers with low scores see previously-promoted workers with similar scores to themselves. Otherwise, low-scoring workers would likely see only top-scoring workers being promoted and would be unlikely to attribute any decision to discrimination. 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. Evaluating every worker under all three procedures generates the counterfactual data on whether workers would have been promoted under any random treatment assignment.

Workers that are not promoted under their randomly-assigned promotion procedure (92.5 percent, N=2,886) are offered the easier job and the 80 percent who take it up become the experimental sample (N=2,397).[16] They are all offered the same Prolific task (i.e. same wage and job description) and only learn about their evaluation *after* opting in to the job. Attrition after starting the survey was uncommon, and similar across all treatment arms (see Section 4.3). Promoted workers are offered the harder job, but are not the sample of interest.

Finally, 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).

Information about non-promotion and evaluation procedures. The timeline of the experimental survey is shown in Figure 4. Upon starting the survey, workers are told about the two jobs, including the first mention of a manager or algorithm. Then, they learn 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*

difference between workers' performance and their (incentivized) beliefs about performance is a measure of confidence, which may have been an important predictor of perceived discrimination if motivated reasoning causes some workers to attribute rejection to discrimination rather than their own performance (Heidhues et al., 2019). In fact, this does not seem to be the case in this setting (Appendix Figure A35).

[16] Those who opt in to the experimental survey are different from those who do not, but not differentially by treatment arm (unsurprisingly, since there had been no differential treatment at this stage).

job?," they then learn how they were evaluated.

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 A2 shows the implementation of each manager arm.

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

In the demographic-blind manager arm, the profiles show only quiz-score quintiles as 1-5 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 sample, all workers learn that a computer 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 A3 shows the implementation. Thus, "treatment" in the algorithm sample 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.

[17] Only 6 of 2397 participants identified as non-binary. Non-binary participants are grouped with women in results that split by gender as the "not male" category. Self-identified races include Asian (13 percent); Black, Hispanic (<1 percent); Black, Non-Hispanic (13 percent); Multiracial, Hispanic (1 percent); Multiracial, Non-Hispanic (<1 percent); Native American, Hispanic (<1 percent); Native American, Non-Hispanic (<1 percent); Other, Hispanic (3 percent); Other, Non-Hispanic (<1 percent); White, Hispanic (10 percent); White, Non-Hispanic (58 percent).

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.[18]

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.[19]

To my knowledge, no other study has attempted to identify individuals who felt that they had been discriminated against in a recent interaction.[20] 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 A4).

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. They are reminded that this will not affect their payment on the already-proofread paragraphs. 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. I observe the time spent on each paragraph, the number of times they click on each paragraph, and how many of their final highlights correctly identify mistakes or incorrectly identify non-mistakes. A summary measure of performance is their earned bonus.

[18] One concern is that the fraction of white men among previously-promoted workers could co-vary with manager characteristics, but this is not reflected in the data and there was minimal heterogeneity in manager profiles. Neither this or any other characteristics of the previously-promoted workers are correlated with what workers know about their manager (Appendix Table A14). Workers who see that two versus three white men were previously promoted have similar observable characteristics (Appendix Table A15).

[19] Note that age was not communicated directly but was somewhat observable through avatar hair color and baldness. This measure 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; cases in which they disagreed were resolved by a PhD psychologist who professionally codes qualitative data and was hired on Upwork. In the hiring experiment, the text was coded once by the same PhD psychologist. In both cases, 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.

[20] Surveys (e.g. Kessler et al., 1999; Goldsmith et al., 2004; Mukerjee, 2014; 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.

After the first, third, or fifth paragraph—timing randomly assigned—workers also complete a standard psychological scale that elicits current emotional states (affective well-being). This is the only self-reported outcome collected during the proofreading job.

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, where 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 for that wage schedule are implemented for a random subset of workers.

Workers answer these questions under two conditions: first, if they would be evaluated by the same procedure as in the experiment and second, if they would be promoted if they had some of the highest baseline quiz scores and otherwise offered the easier job again. After each multiple price list, they report what they think is the probability that they will be promoted if they participate in the future round with the given evaluation procedure. Workers knew there may be future job opportunities from the beginning of the survey, but did not know that they may interact again with their manager until the reservation wage elicitation.

In the manager arms only, I similarly elicit reservation wages to participate in a collaborative and interactive task with their manager from the experiment and their willingness to pay to be able to choose which manager to work with in the cooperative task.[21] 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, which could be reflected in workers' willingness to stay late at work, take on extra tasks, etc. for their manager, all of which may be important for their career trajectory.

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

Secondary measures of perceived discrimination. Finally, they answer a series of questions that also measure perceived discrimination. First, if they say they have a complaint about the promotion procedure 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

[21] 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.

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. Validating my main measure, these more direct measures of perceived discrimination are highly correlated with whether workers mention any of their demographics when asked what needed to be different about their profile for them to be promoted (Appendix Figure A4). 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.[22]

2.3 The hiring experiment

The hiring experiment disentangles the effects of using an algorithm versus a manager and the effect of blinding decision-makers to race and gender when workers see that primarily white men were previously hired. It occurs in the same dynamic labor market on Prolific as the promotion experiment. The not-hired workers take an experimental survey without any proofreading job. The key outcome is perceived discrimination, and the treatment arms separate the effects of (i) manager versus algorithmic evaluation (ii) learning that previous hires are primarily white men and (iii) allowing or disallowing decision-makers to use demographic information, using a factorial design.

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 (Figure 5). 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 the previous hires or not. Again, workers are evaluated by all four procedures to generate counterfactual decisions, 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. Managers evaluate nine groups of 40 workers and their decisions are implemented for randomly-selected groups—a scaled-up equivalent to the promotion experiment.

The demographic-blind algorithm is the same as in the promotion experiment. The non-blind

[22] 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 (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. I discuss these further in Appendix I.

algorithm interacts each baseline performance and education variable in the demographic-blind algorithm with the eight race × 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.[23] They then answer the question about what needed to be different about their profile in order to be hired, which underlies the main measure of perceived discrimination. They do *not* subsequently do a proofreading job, but continue directly to the reservation wage elicitations—now over the opportunity to be evaluated again and potentially hired in the future. Finally, they answer questions that measure their memory and comprehension of the evaluation procedure that determined they would not be hired and those that underly the secondary measures of perceived discrimination.

2.4 Experimental fidelity: Attention and comprehension

The primary mechanism ensuring attention to the information about evaluation procedures was the open-ended question underlying the main measure of perceived discrimination, *"what do you think needed to be different about your profile to be assigned to the harder task?"* All 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 identify 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, and this did not vary by treatment group.

[23] The only difference between the promotion and hiring experiments in treatment implementation is that in addition to seeing a written description of the information known by decision-makers, workers also see their own profile showing the information the decision-maker used before seeing the previous hires' profiles that either have avatars, race, and gender or not. This ensures that perceptions of discrimination in the demographic-blind decision-maker arms where workers see previous workers' avatars, race, and gender are not driven by differential attention to text-based versus visual information as could have been the case in the promotion experiment (Appendix Figure A36).

In both studies, workers answered an "attention check" in the middle of the final survey questions. About 75 and 80 percent of workers answered the question correctly in the promotion and hiring experiments, respectively.[24] Note that the proofreading job and reservation wage elicitations occur before this question was asked and were clearly incentivized, as opposed to the section of survey questions. There is no differential attention across treatment arms and the main results are robust to dropping the workers who failed the attention check (Section 4).

2.5 Specifications

I focus on intent-to-treat (ITT) specifications that compare outcomes between workers assigned to different treatment arms. First, in the promotion experiment, I restrict the sample to those evaluated by a manager or those evaluated by the algorithm and estimate the following:

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

In the sample of workers evaluated by a manager, T_i is one if worker i is in the non-blind manager arm and zero otherwise. In the sample of workers evaluated by the algorithm, T_i is one if worker i sees previously-promoted workers' avatars, race, and gender and zero otherwise. β is the coefficient of interest. In the manager sample, β is the effect of learning that one's manager saw avatars, race, and gender and previously promoted mostly white men. In the algorithm sample, β is the effect of learning that the demographic-blind algorithm previously promoted mostly white men. In both samples, γ_g is fixed effects for workers' quiz-score quintile group, X is a vector of baseline worker characteristics and H is a vector describing the education and quiz scores of the previously-promoted workers seen by worker i.[25] Appendix D provides details on the construction of all outcome and control variables. Standard errors are robust to heteroskedasticity.

In the hiring experiment, 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 evaluated by an algorithm rather than a manager, BA_i is an indicator for being evaluated by a demographic-blind decision-maker and seeing previous hires'

[24] The question asked workers to select one specific column for both rows in a matrix with two Likert-scale style questions and alerted them that it was in order to make sure they were paying attention.

[25] 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 seven indicators for whether they saw one, two, or three workers with more or less education than themself or whether they saw three workers with the same education levels as themself. The controls in H were not pre-registered, because I did not anticipate that different types of promotion procedures would make meaningfully different choices of workers in terms of their education and previous performance. The results are the same when I exclude these controls (Section 4).

avatars, race, and gender, NB_i is an indicator for being evaluated by a non-blind decision-maker and seeing previous hires' avatars, race, and gender, and all other variables are the same as above. 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 when the decision-maker is demographic-blind and workers do know that previous hires were mostly white men, and χ_2 when the decision-makers do know race and gender and workers know that previous hires were mostly white men. The effect of learning that a demographic-blind manager previously hired mostly white men is θ_1 and the effect of learning that a manager knew race and gender is $\theta_2 - \theta_1$. The effect of learning that a demographic-blind algorithm previously hired mostly white men is $\theta_1 + \chi_1$ and the effect of learning that an algorithm used race and gender is $\theta_2 - \theta_1 + \chi_2 - \chi_1$.

One threat to the causal interpretation of these specifications as the effects of interventions that change only *perceptions* of discrimination is that different evaluation procedures could have selected different types of workers. Since their real decisions were implemented, this could generate differences between the treatment arms among workers who are not selected. To eliminate this threat, I built into the experimental design 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).

No decisions described in this section affect the main results: conclusions are the same when I include different controls (no controls, lasso-selected controls), or either expand to the full sample or further restrict the sample to workers who passed an attention check, as well as when I use randomization inference or adjust for multiple hypothesis testing (Section 4).

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 by the effect of the treatments 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. Of course, 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 Results

3.1 Perceived discrimination

Workers' perception of discrimination strongly depends on what they learn about evaluation procedures. First, I'll discuss treatment effects on perceived discrimination in the promotion experiment. These are most useful for understanding the implicit "first stage" when turning to the results on worker behavior in Sections 3.2 and 3.3. Then, I turn to understanding the effects of procedural changes (from the hiring experiment) and minority representation (from the secondary randomization in the promotion experiment) on perceptions of discrimination.

Treatment effects in the promotion experiment. Figure 6 plots rates of perceived discrimination in each arm overall and separately for women, racial minority men, and white men. Table 2 presents the analogous treatment effects using the specification described above.[26]

Perceived discrimination is low among workers in the demographic-blind manager arm. About 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 few workers hypothesize that the manager had more information than was shown in the profiles, including race and gender (which makes sense since workers saw the race and gender of the manager).

In contrast, 40 percent of women and 20 percent of racial minority men perceive discrimination in the non-blind manager arm (Figure 6). Six white men (7 percent) also perceive discrimination—three of them reference age discrimination (communicated by gray hair in their avatars), and the other three seem confused about the promotion process but mention demographics. Overall, being in the non-blind manager arm increases the share of workers perceiving discrimination by 31pp (se=2pp) (Table 2, Panel A).[27]

Similar patterns emerge when comparing workers in the algorithm arm who either saw that the

[26] Appendix Figure A37 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. Heterogeneity by race and gender was pre-registered, along with heterogeneity by whether workers have past experiences of discrimination, high or low beliefs about the prevalence of discrimination, and high or low confidence. I focus on gender and race as they most systematically predict perceived discrimination, but multivariate regressions that include all pre-registered characteristics are in Appendix Figure A35.

[27] Almost all workers who do not perceive discrimination cite needing more education or higher quiz scores to be promoted. Being in the non-blind manager arm reduces the share who cite the *primary* reason being their education or quiz scores by 10pp and 6pp (25 and 13 percent, se=2.4pp and 2.6pp), respectively (Appendix Table A16). 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).

algorithm previously promoted mostly white men or not. 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. 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." When workers know that the algorithm previously promoted mostly white men, however, many perceive discrimination: 22 percent of women and 10 percent of racial minority men (Figure 6). Overall, learning that the algorithm previously promoted mostly white men increases the share of workers perceiving discrimination from 0 to 17pp (se=2.5pp), even though they are told that the algorithm did not use demographics to select them (Table 2, Panel B).[28]

Appendix Tables A5 and A6 show parallel results using the secondary measures of perceived discrimination. There is always a large, significant effect in the full sample, and patterns across demographic groups are the same for all measures, though the prevalence in the control groups and magnitude of the treatment effects vary.

The effect of de-biasing hiring procedures. In the promotion experiment, subjects in the non-blind manager arm learn two things compared to the blind manager arm: the characteristics of previously-promoted workers, and the fact that the manager knew and potentially used their demographics. Both could influence perceived discrimination. Additionally, differences in the rates of perceived discrimination in the manager and algorithm samples could come either from the fact that the algorithm does not use demographic information or that it is an algorithm. To disentangle these mechanisms, I designed and implemented the hiring experiment. I focus on whether procedural changes that reduce the likelihood of actual discrimination—blinding managers to demographics and using unbiased algorithms—mitigate perceptions of discrimination.

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 setting when minority-groups are severely under-represented. In the hiring experiment, almost all workers saw exclusively white men being hired. In the main analysis, I restrict to those workers (a random subset, as in the promotion experiment). Results for the full sample are very similar (Appendix Figure A5).[29]

[28] 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 A17) while it increases the share who cite discrimination as the *primary* reason by 9pp (se=2pp, Appendix Table A6).

[29] The different 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).

All else equal and *when workers see only white men were previously promoted*, changing hiring procedures in a way that reduces or eliminates the scope for human bias to affect decisions cannot eliminate perceptions of discrimination. Figure 7 plots rates of perceived discrimination in each treatment arm in the hiring experiment estimated using the main specification. Like the promotion experiment, no workers perceive discrimination in the demographic-blind manager or algorithm arms where workers do not see previously-promoted workers' avatars, race, and gender, so I focus on the other arms.

Workers perceive *more* discrimination by algorithms than managers. When decision-makers see workers' avatars, race, and gender and workers know all previous hires were white men, 52 percent of workers evaluated by a manager 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 7).

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, though magnitudes vary (Appendix Figure A6).

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. This is *not* due to (differential) inattention or general confusion about the hiring procedures or experimental design (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 A7). 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 from differential forgetfulness or confusion about this particular decision input, for example because they saw two different types of profiles instead of one.[30]

The results suggest that at least some workers understand that algorithms can discriminate

[30]The information about what the decision-maker knew and the information about who was previously promoted should have been similarly salient as both were communicated visually (Appendix Figure A36).

even if they only use 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. This 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 decision-makers used demographic information between the manager arms and the algorithm arms (Appendix Figure A7). Recall that workers are 9pp (21 percent) *more* likely to perceive discrimination by an algorithm than a manager when both are demographic-blind but workers know that previous hires are all white men. Together, these results suggest that some workers know that 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. 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 pay structure and incentives. So, misunderstanding the algorithm may be another reason they are more likely to perceive algorithmic than manager discrimination (Appendix Figure A8).

The effects of minority-group representation. On the other hand, more representation of women and racial minority men among previously-selected workers substantially reduces perceived discrimination. In the promotion experiment there is random variation in whether workers saw three white men previously promoted versus two white men and someone else. Seeing one minority-group member among the three previously-promoted workers reduces rates of perceived discrimination from 42 to 25 percent in the non-blind manager arm (a 17pp difference, se=3pp) and from 30 to 16 percent in the algorithm arm when workers see previously-promoted workers' demographics (a 14pp difference, se=5pp) (Figure 8). In pilot data, when workers do not see any information about managers' past decisions, learning that a manager knew race and gender when making a decision only increases perceived discrimination from 0 to 7 percent. Thus, workers from oft-discriminated groups do not perceive discrimination without some additional signal that it occurred but when a situation "looks like discrimination," they are likely to perceive that it was so. In other words, their beliefs 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.

Taken together, these results have several implications for settings where some demographic groups are severely under-represented. First, anti-discrimination policies—like transitioning to algorithmic screening or censoring demographic identifiers from resumes before they are reviewed—may slightly reduce perceptions of discrimination on their own, but if there is no change in the demographics of workers who are selected, workers may still be likely to perceive that discrimination played a role. On the other hand, when workplaces are more diverse—which could be obtained by these types of anti-discrimination policies—workers will be less likely to see decisions that "look

like" discrimination, and this may substantially reduce perceptions of discrimination.

3.2 Retention and performance

The promotion experiment measures the effects of perceived discrimination on retention and performance. Workers complete a proofreading job after learning how they were evaluated and not promoted. 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. Learning that a demographic-blind algorithm previously promoted mostly white men lowers performance and earnings.

Retention. Workers must proofread six paragraphs to be paid their participation wage, and can then choose to proofread up to eighteen. Recall that 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 by the main measure. It also causes workers to proofread 0.48 fewer paragraphs (3 percent, se=0.25, p=0.05) and reduces the probability that they proofread all eighteen paragraphs by 4.5pp (6 percent, se=2.5pp, p=0.07) (Table 3, Panel A). On the other hand, learning that the demographic-blind algorithm previously promoted mostly white men—which increases perceived discrimination by 17pp—does not affect retention (Panel B).

Intensive effort and performance. Perceived manager and algorithmic discrimination both worsen performance. I identify causal effects of treatment on 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 causes workers to incorrectly highlight 0.36 more non-mistakes (11 percent, se=0.22, p=0.10), and learning that the algorithm previously promoted mostly white men causes workers to correctly highlight 0.7 fewer mistakes (5 percent, se=0.38, p=0.07) and earn 6 cents less (6 percent, se=3.6 cents, p=0.10) (Table 4). However, there are no effects on measures of how hard workers choose to work: how much time to spend per paragraph (capped at 1 minute) or the number of times they click on the page (a proxy for whether they check their answers).

Total earnings. If the bonus was a piece-rate paid per paragraph, the negative retention effect just described would correspond directly 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 a treatment 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, and 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 (Appendix Figure A9).

Heterogeneity by race and gender. The negative effects on retention are driven by women, and racial minority women drive the negative performance effects. In contrast, racial minority men experience positive performance effects.[31] Figure 9 plots the effect of being in the non-blind manager arm rather than the demographic-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 4.1pp (73 percent, se=2.1pp, p=0.05) 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 0.79 *more* paragraphs, though the estimate is noisy and not significant. The effects on retention for men and women are significantly different from each other at the 5-10 percent level depending on the outcome (Appendix Table A7, Panel A. Panel B pools white and racial minority women and shows 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 0.9 (33 percent, se=0.39) more non-mistakes (Appendix Table A8, 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=0.05*)* from the zero effect for women when white and racial minority women are pooled together (Panel B).[32]

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—

[31] I focus on heterogeneity in the manager sample because the algorithm sample is half the size.

[32] 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 9). The negative point estimate of the effect on women's total earnings is likely due to the large negative effect on their retention. The fact that it is less stark than the effects on retention has two possible explanations that I cannot separate: (i) the women induced to leave are negatively selected and would not have earned much in the absence of discrimination, and (ii) there is a positive treatment effect on those who stay. The overall positive effect on racial minority men's total earnings likely comes from positive treatment effects, given the positive effect in the first six paragraphs and little evidence of differential retention.

they are more annoyed and less motivated, though also less anxious (Appendix Figure A10).[33]

In summary, perceived discrimination—by managers or algorithms—overall has negative effects on performance, retention, and earnings. Some workers may react by trying to prove themselves to biased managers, but only some do so successfully. While the gender heterogeneity identified here may arise partly because gender discrimination was more salient in this context, it suggests that in cases like this, perceived discrimination may particularly entrench gender gaps because women are more negatively affected—particularly among racial minorities.[34]

3.3 Future labor supply

Consistent with the effects on retention, perceived discrimination reduces workers' future labor supply when they anticipate that they will face additional discrimination by a manager or an algorithm. These effects can be partially explained by beliefs about lower expected wages but derive primarily from anticipated disutility from discrimination or interactions with a biased manager in the future. They are measured in both the promotion and hiring experiments, discussed in turn.

3.3.1 Promotion experiment

After workers complete the proofreading job, I remind them of how they were assigned, emphasize that some workers will be offered additional work opportunities in the future, and elicit reservation wages for these future 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 always twice the lower piece rate. 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 5 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.[35] I use a 50 percent probability because the average worker in the control groups believes this is their chance of being promoted in the future.[36]

[33] There are no effects on self-efficacy or job satisfaction (Appendix Tables A18 and A19)

[34] 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 the other dimensions of pre-registered 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.

[35] Sample sizes are slightly smaller for these variables as I drop observations in which workers make inconsistent choices that indicate they do not understand the multiple price list elicitation method (e.g. that they would take the job if it paid 0.50 per paragraph but not if it paid $0.55 per paragraph), which is standard in this type of analysis.

[36] This construction puts this overall reservation wage on the same scale as the variable that accounts for workers' beliefs about future promotion in the next section. Note that 50 percent is much higher than the actual promotion rate, but workers had no information about the actual promotion rate. Appendix Figure A38 plots the CDFs of beliefs by treatment group in both experiments and shows that a large share of workers guess 50 percent, suggesting flat

Thus, the effect on this formulation of reservation wages could be driven either by beliefs about future discrimination (and thus promotion) or psychological mechanisms like anticipated disutility from future discrimination. I decompose it into these two channels in Section 3.3.2.

Perceived discrimination increases workers' reservation wages for future work and their willingness to pay for a transparently unbiased mechanism. Being in the non-blind manager arm increases reservation wages from 26.5 cents to 29 cents per paragraph (9 percent, se=1 cent) if workers would be evaluated by the same manager again in the future (Table 5, column 1). 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 from the main experiment (column 2), measured as the difference in reservation wages if they would be evaluated by their manager again or the cutoff rule. This willingness to pay is economically meaningful: if the future wage is the same as in the experiment (25 cents per paragraph, also approximately the average reservation wage in the demographic-blind manager arm), the increased willingness to pay for the cutoff caused by the treatment is 4 percent of their expected future payment if they anticipate not being promoted. Learning that the demographic-blind algorithm previously promoted mostly white men has no significant effect on reservation wages or willingness to pay for the unbiased cutoff in the promotion experiment, but the effects are not statistically distinguishable from the positive effects in the manager arms (Table 5, panel B).

Consistently, perceived discrimination also reduces workers' willingness to interact with and generosity towards their manager. Table 6 shows that 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 4.3 cents per paragraph (20 percent, se=2 cents). Workers in the non-blind manager arm are 5pp (10 percent, se=3pp, p=0.08) more likely to share \$0 with their manager of their possible \$20 bonus, and overall share about 30 cents (11 percent, se=23 cents, p-value=0.19) less with their manager (Table 6). This coefficient is quite stable across various specifications and is significantly different from zero at the 5-10 percent confidence level in some specifications (see Section 4).[37]

Applied to a work context where employees and managers interact on a regular basis, this response to perceived discrimination could be self-fulfilling—if workers react to perceived dis-

priors, but that there is wide variation. Reassuringly, workers with higher baseline confidence have higher beliefs about the likelihood of promotion (not shown).

[37]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 of dictator games suggests that workers would have shared around \$6 (30 percent) with strangers (Engel, 2011), so we can infer that even workers in the demographic-blind manager arm may be retaliating against their managers for not promoting them in the first place or that the hierarchical nature of the manager-employee framing affects distributional preferences. The effect of being in the non-blind manager arm is the effect on top of any effect of non-promotion or the manager-employee framing.

crimination by avoiding their managers or refusing their requests, managers may be less likely to reward these workers in the future—which may itself be perceived as discrimination. Managers may also react more negatively to women or racial minority men who under-perform in this way, in which case these effects may be even more consequential. While I do not detect evidence of heterogenous treatment effects on future labor supply, these effects may still exacerbate gaps if managers react more negatively to these changes in behavior for members of certain groups.[38]

3.3.2 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. These effects, 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 (Lundberg and Startz, 1983; Coate and Loury, 1993; Fryer et al., 2005; de Haan et al., 2017; Dianat et al., 2020)—or does not separate the two (Alston, 2019; Charness et al., 2020; Ridley, 2022; Aksoy et al., 2023; Avery et al., 2023). Suggesting that the first may be important, Dube et al. (2022) and Folke and Rickne (2022) show that workers value similar non-monetary amenities like dignity at work or a lack of sexual harassment.

Understanding how each channel contributes to the effects of perceived discrimination clarifies whether these effects are commensurate with other changes in expected wages, or instead, if perceived discrimination may be especially pernicious and demand particular attention. Further, it 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 being promoted in the future. Specifically, 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.

[38] In terms of race and gender, being in the non-bind manager arm increases all workers' willingness to pay to avoid their manager or discrimination in the future, via either reservation wages, willingness to pay for the unbiased cutoff rule, or willingness to pay for a new manager (Appendix Table A9). 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.

Let D be perceived discrimination in a previous encounter and $\hat{\sigma}$ be beliefs about the likelihood of future discrimination. In the standard model above, 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 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 for this future promotion opportunity.

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 $\hat{\sigma}$: $\hat{\pi} = \pi(\hat{\sigma}, D)$. Intuitively, perceived discrimination may change e.g. self-confidence or trust in the employer, which affects beliefs about promotion independently of anticipated discrimination. Second, workers may anticipate a cost $C(\hat{\sigma}, D)$ which can affect expected utility independently of $\hat{\pi}$. Intuitively, workers may dislike discrimination—unfairness generally, or specifically being treated unfairly because of one's race or gender—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:

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

The effect of perceived discrimination that operates through $C(\cdot)$ may be particularly large if being discriminated against again feels worse when one has just experienced discrimination.

Results. The evidence from my setting rejects both implications of the standard model. First, I'll show that 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 30 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 in general.

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 $\hat{\sigma} = 0$.[39]

[39] This presumes that workers believe what I say about the procedures that will be used in the future, which I consider part of $C(\cdot)$ (the trust channel described above).

I find that being in the non-blind manager arm does increase anticipated discrimination, but also 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 3pp (6 percent, se=1.1pp). Half of this effect, however, can be explained by changes in beliefs about the likelihood of promotion when no future discrimination is possible (i.e., $\hat{\sigma} = 0$) (Table 7, columns 1 and 3 in Panel A). Thus, half of the effect of perceived discrimination on beliefs about the likelihood of future promotion comes from increased anticipated discrimination.[40]

Next, I estimate the effects of perceived discrimination on reservation wages, constructed in a way that accounts for changes in beliefs about the likelihood of promotion. Any effect on this construction of reservation wages isolates the role of $A(\hat{\sigma}, D)$. Recall that the reservation wages described earlier (Table 5) 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.[41] 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 $\frac{\partial C}{\partial D}$.

Learning that their manager knew race and gender and previously promoted mostly white men 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=0.09), using this construction of reservation wages. This is 70 percent of the total effect on reservation wages (2.4 cents per paragraph) when beliefs about promotion can also play a role (Figure 10, top panel). When workers will be evaluated by the unbiased cutoff rule, treatment increases reservation wages by 0.9 cents per paragraph after accounting for beliefs (though the effect is not statistically significantly different from zero). This effect derives only from disutility from working in a place where one has experienced discrimination in the past. The overall effect of 1.7 cents combines this effect with the effect of anticipating interacting with the perceived racist or sexist manager or experiencing additional discrimination. Taking the point estimates at face value, this means that about half of the overall effect is therefore due to disutility from anticipated discrimination or continued interactions, and the other half is due

[40] Learning that the algorithm previously promoted mostly white men does not increase anticipated discrimination in the promotion experiment, but does in the hiring experiment (discussed next, along with possible explanations for the difference).

[41] 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.

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 7.5pp (11 percent, se=2.7pp), are interested in more jobs with tasks assigned like this one by 4.6pp (8 percent, se=2.8pp, p=0.10), and are interested in doing the harder job in the future by 6pp (11 percent, se=2.8pp) (Appendix Table A10).

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 types of preferences 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.[42] Thus, it seems 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.

3.3.3 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.[43] In Panel A of Table 8, 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 A11). Column 1 shows the effect of treatment on perceived discrimination in the pooled treatments (36pp

[42] This was intentionally left ambiguous, but workers did not think this was the case. Towards the end of the survey, workers were asked the open-ended question, *"how do you think it is determined whether you did a "good job" [and thus are paid for] proofreading these paragraphs?"* Fewer than 1.5 percent of workers in either manager arm mentioned "manager" 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 some set of predetermined answers to compare to, and many workers also mentioned a cutoff rule in the number of correct answers.

[43] 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.

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 8, column 2). 88 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 8, 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 (Appendix Table A11).

These results are consistent with other evidence that workers are willing to pay to avoid discrimination (Alston, 2019) and that anticipated discrimination reduces labor supply (Avery et al., 2023). In addition, I document for the first time that perceived and anticipated discrimination cause direct disutility, thus affecting labor supply even after accounting for the monetary consequences of anticipated discrimination or when there is no possibility of future discrimination.

4 Robustness

This section discusses the stability of the results when accounting for standard concerns with experiments in general and survey experiments in particular. The paper's conclusions are unaffected.

4.1 Randomization inference and multiple hypothesis testing

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.

Appendix Figure A12 plots the p-values of the differences in rates of perceived discrimination in different arms of the hiring experiment alongside randomization inference p-values. Appendix Figure A13 does the same for the effects of assignment by a non-blind manager in the promotion experiment. In both cases, conclusions are the same.

Appendix Figures A12 and A13 also plot the corresponding p-values when they are corrected for multiple hypothesis testing. Specifically, I control the family-wise error rate using the Romano-Wolf (2005; 2016) correction with 500 bootstrap iterations. In Appendix Figure A12 the four estimates are considered one "family," and in Appendix Figure A13 the four "families" are the groups in bold: retention, effort and performance in the first six paragraphs, future labor supply, and beliefs about future promotion. Again, inference is generally unaffected, though the effects of being in the non-blind manager arm on retention now have p-values between 0.1 and 0.15.

4.2 Specification choices

I test the sensitivity of the main estimates to the choice of control variables and sample definition. Appendix Figures A14 and A15 plot the share of workers perceiving discrimination in each of the four arms of the hiring experiment with positive perceived discrimination under each set of decisions. Appendix Figures A16-A23 (manager sample, promotion experiment) and A24-A28 (algorithm sample, promotion experiment) each correspond to a set of outcomes. Each figure starts with the main specification and iterates through the combinations of each decision point, plotting the estimate of each treatment effect when that combination of decisions is used.

4.2.1 Sample definitions

The main results restrict the sample to workers who were not selected by any of the procedures, as pre-registered. This eliminates concerns that different procedures may have selected different types

of workers, generating differences between the treatment arms in the non-promoted or non-hired sample other than those caused by different perceptions of the reason they were not selected.

The specification charts (Appendix Figures A14-A28) 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 choice of sample generally does not change the estimated treatment effects, though 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 A27). 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 A18).

4.2.2 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). 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. The figures also show estimates from regressions that include no control variables.

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

[44] All controls in the main specification, plus: indicators for reporting having experienced gender or racial discrimi-

by an algorithm and separately for each sample definition in the previous subsection. For each outcome × 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.

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. Unsurprisingly, adding evaluation-group fixed effects tends to improve the precision of the estimated coefficient, and removing all controls does the opposite. Overall, the conclusions are unchanged.

4.3 Attrition and attention

Appendix Table A12 shows that there is no differential attrition in the promotion experiment between the demographic-blind and non-blind manager arms, and minimal attrition between workers in the algorithm sample that do or do not see avatars, race, and gender. In the manager sample, there is no difference in the rate at which workers started the experiment after completing the baseline survey, finished the experiment if they started it, or failed the survey attention check. While workers who failed the attention check are included in the main sample (results are robust to dropping them, see above), this provides reassuring evidence that the results are not driven by differential attention across treatment arms. Section 2.4 discusses overall experimental fidelity and shows that worker attention to and comprehension of the experiment was high overall.

The only case of differential attrition or attention in the promotion experiment is that in the algorithm sample, workers who see the avatars, race, and gender of previously-promoted workers are 3.3pp (3 percent, se=1.8pp) more likely to finish the experiment if they started it than those who do not see demographic information. The large difference in perceived discrimination (17pp, se=2.5pp) between these workers cannot be explained by this amount of differential attrition, particularly as it was measured at the beginning of the survey. None of the other main conclusions of the paper hinge on the algorithm arm of the promotion experiment.

nation in past job search, promotion, termination, or daily work activities, second-order beliefs about those probabilities for each of eight race × 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 × gender × quintiles × degree status as shown on the profiles.

There is no differential attrition across treatment arms in the hiring experiment or differential rates of passing the attention check (Appendix Figure A29).

4.4 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]

Unsurprisingly, Appendix Table A13 (promotion experiment) and Appendix Figure A30 (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 managers or an algorithm, respectively. This is likely due to the final questions in the survey that ask about perceptions of discrimination most explicitly.

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

[45] Mummolo and Peterson (2019) show that giving participants information about experimenter hypotheses does not affect participant behavior or estimated treatment effects, even when participants are financially incentivized to conform to those hypotheses and participants are 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.

positive rates of perceived discrimination (Appendix Figure A30). 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. In the promotion experiment, I estimate conservative bounds on the effects on worker behavior 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.[47] I use knowledge of the study topic *as measured at the end of the survey*, a conservative measure given what smaller fractions of workers guessed the study topic at the time the main outcomes were measured. Using the measure of known study purpose that was elicited as quickly after that outcome as possible leads to similar conclusions but restricts the analysis to half the sample. 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. The estimated treatment effects are similar and my conclusions are unchanged (Appendix Figures A31 and A32).

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. The estimated differences in rates of perceived discrimination between treatment arms are similar (Appendix Figure A33)

4.5 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—see Appendix G for more detail.

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 groups of historical workers who were jointly evaluated. I did not anticipate the amount of variation there would be in composition of previously-promoted workers. Given that (i) there was substantial variation and (ii) the effect of

[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.

minority-group representation on perceived discrimination I document 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, there was a change in the promotion experiment's demographic composition, which originally planned to more substantially oversample 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.

5 Conclusion

This paper shows that addressing the disparate impacts of discrimination will require addressing worker perceptions in addition to discrimination itself. Perceived discrimination has large, negative effects on 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 underrepresented. 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 signals that "look like discrimination." Increased diversity seems to be a more promising tool to reduce perceptions of discrimination. These results offer a new efficiency argument for affirmative action or other diversity initiatives (in addition to other benefits, e.g. Rock and Grant (2016); Calder-Wang and Gompers (2021)).

These results also offer interesting avenues for future research on perceptions of discrimination, in that this paper has primarily identified a problem and not yet a solution. Can workers be convinced that algorithms are unbiased? More generally, how can they be encouraged to trust information about de-biased decision-making? On the other hand, diversity initiatives aiming to improve minority representation are often accompanied by efforts to improve equity and inclusion—how do those efforts affect perceptions of discrimination? Relatedly, implicit bias training is not very effective at reducing bias (Chang et al., 2019)—is it so popular because it helps reduce perceptions

of discrimination? 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.*

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

Aksoy, Billur, Ian Chadd, and Boon Han Koh, "Sexual identity, gender, and anticipated discrimination in prosocial behavior," *European Economic Review*, 2023, *154*, 104427.

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

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

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 Expla-

nations," *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.

Burn, Ian, Daniel Firoozi, Daniel Ladd, and David Neumark, "Age Discrimination and Age Stereotypes in Job Ads," 2023.

Calder-Wang, Sophie and Paul A. Gompers, "And the Children Shall Lead: Gender Diversity and Performance in Venture Capital," *Working Paper*, 2021.

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.

Carpio, Lucia Del and Thomas Fujiwara, "Do Gender-Neutral Job Ads Promote Diversity? Experimental Evidence from Latin America's Tech Sector," *NBER Working Paper No. 31314*, 2023.

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.

Hoff, Karla and Priyanka Pandey, "Discrimination, Social Identity, and Durable Inequalities," *American Economic Review*, 2006, *96* (2), 206–211.

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.

Kang, Sonia K., Katherine A. DeCelles, András Tilcsik, and Sora Jun, "Whitened Résumés: Race and Self-Presentation in the Labor Market," *Administrative Science Quarterly*, 2016, *61* (3), 469–502.

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.

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.

Mullainathan, Sendhil, "A Memory-based Model of Bounded Rationality," *The Quarterly Journal of Economics*, 2002, *117* (3), 735–774.

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 Sociology*, 2015, *120* (4), 1005–1054.

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

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.

Rock, David and Heidi Grant, "Why Diverse Teams Are Smarter," *Harvard Business Review*, 2016.

Romano, Joseph P. and Michael Wolf, "Stepwise Multiple Testing as Formalized Data Snoop-

ing," *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.

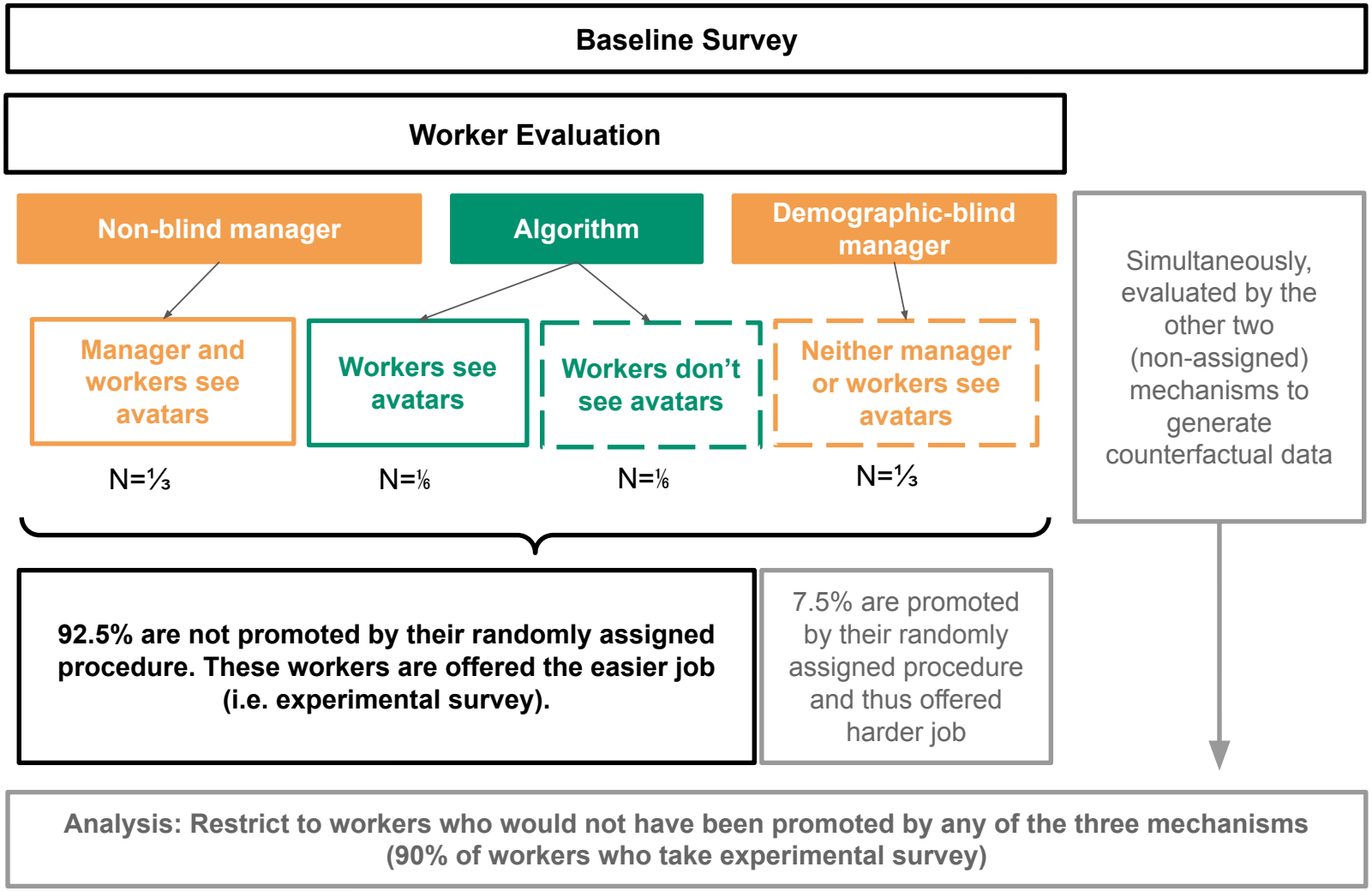
Small, Mario L. and Devah Pager, "Sociological Perspectives on Racial Discrimination," *Journal of Economic Perspectives*, 2020, *34* (2), 49–67.

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

University of Minnesota, "IPUMS USA," 2021.

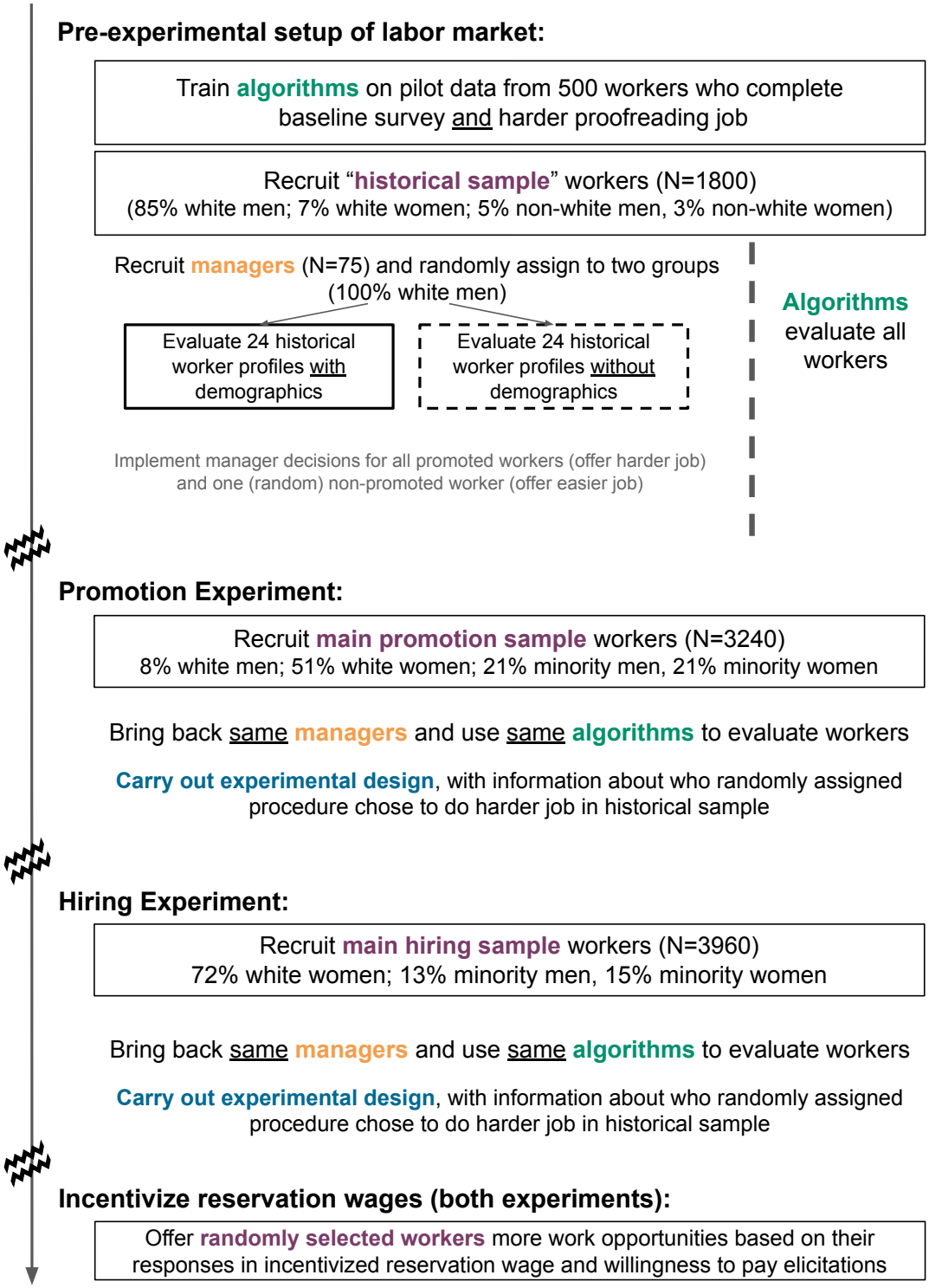
Zhang, Lixuan and Christopher Yencha, "Examining Perceptions Towards Hiring Algorithms," *Technology in Society*, 2022, *68*, 101848.

Figure 1: Experimental design of the promotion experiment



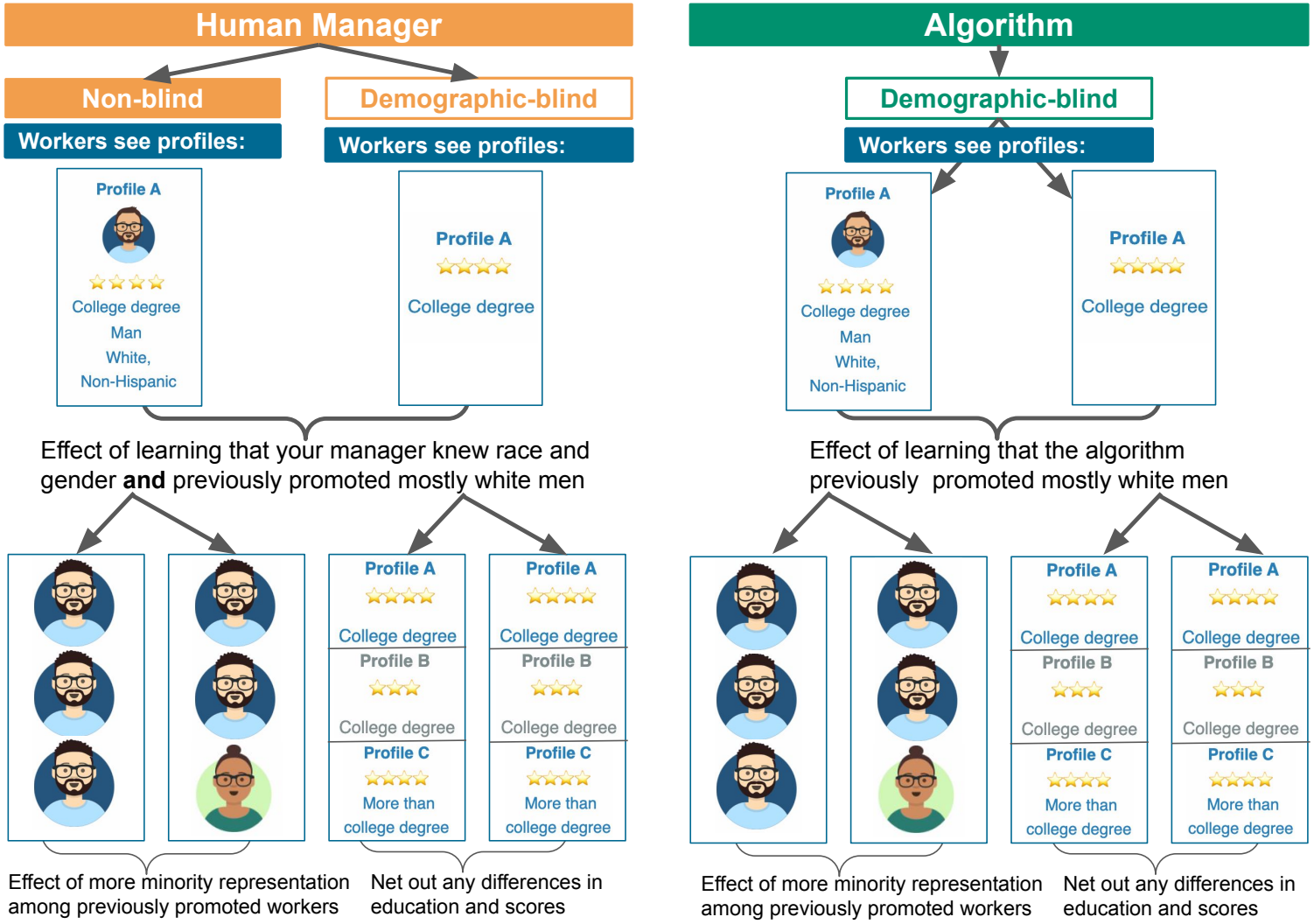
Note: This figure describes the design of the promotion experiment. The baseline survey includes grammar, science, and spelling quizzes and confidence in each, demographics, work history, and past experiences of and beliefs about the prevalence of discrimination. The set-up for workers in the historical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

Figure 2: Timeline of study



Note: This figure describes the dynamics of the labor market created on Prolific to generate repeated interactions with workers and managers. The pre-experimental setup describes the recruitment of the historical sample and managers. The same managers were brought back to evaluate the workers in the experimental samples. The main difference between the historical and experimental samples was the demographic composition, in order to approximate a scenario in which under-represented workers saw that their manager had mostly promoted workers from the majority group in the past. After the conclusion of the experiment, randomly-selected workers are offered additional proofreading jobs that implement the scenarios in the reservation wage and willingness to pay elicitations.
[Click to return to Section 2 (Research Design)]

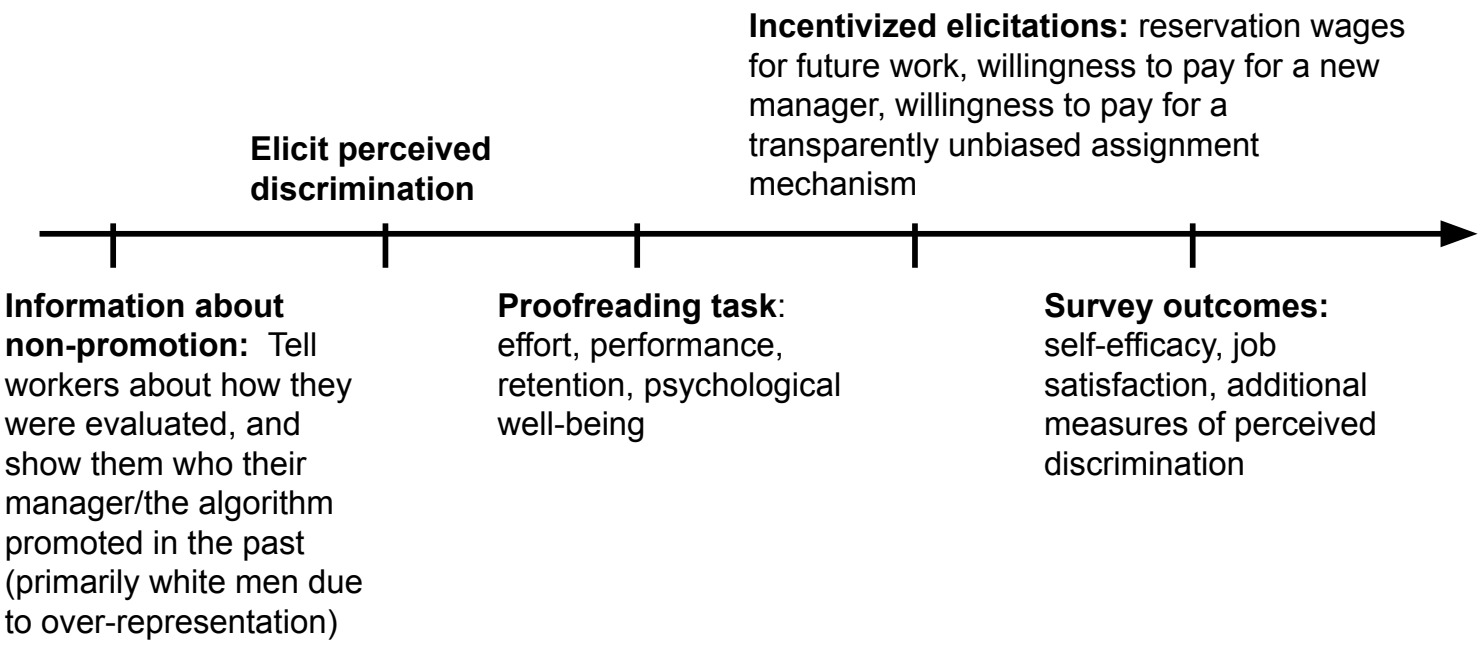
Figure 3: Design of promotion experiment



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.

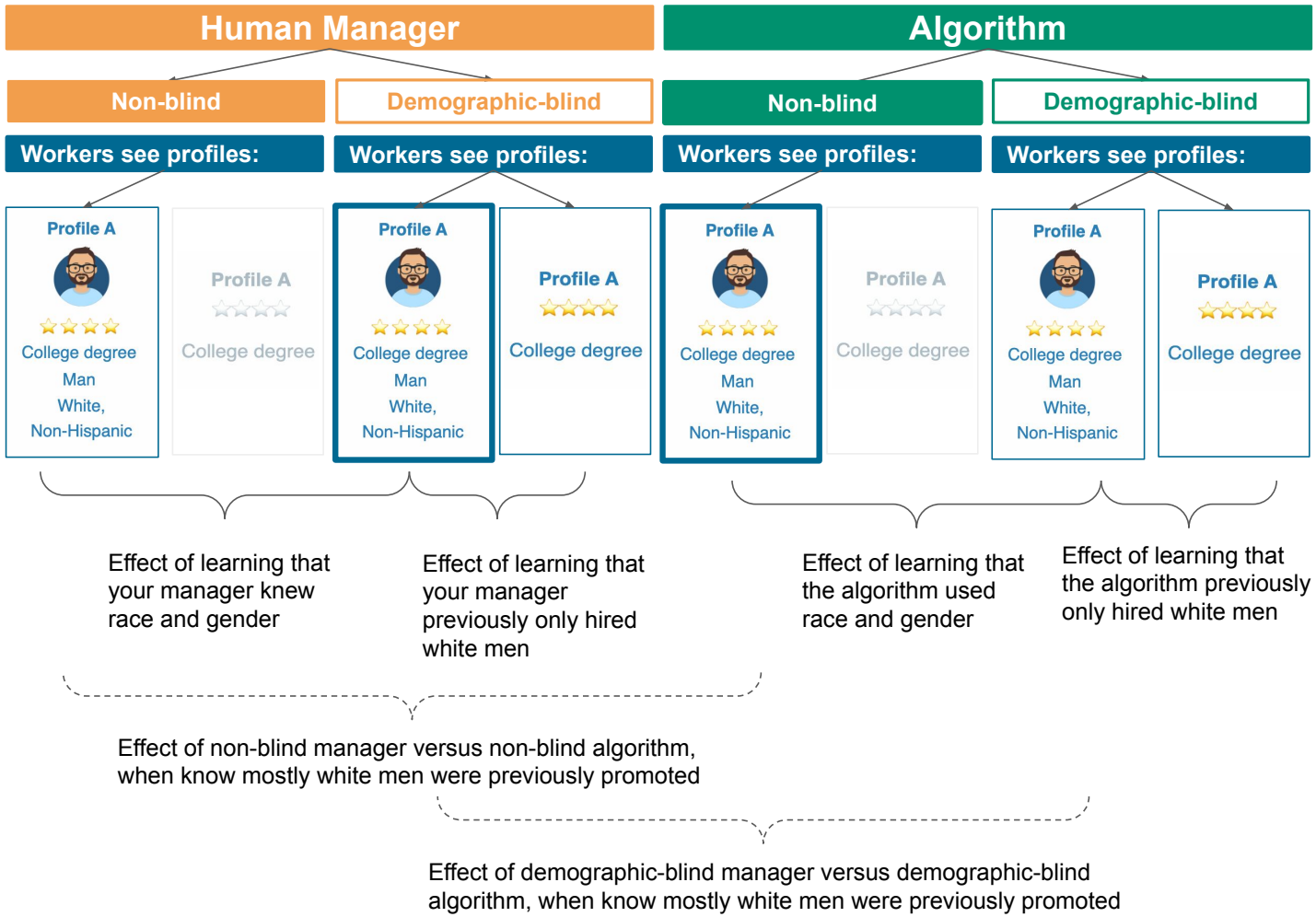
[Click to return to Section 1 (Introduction)]
[Click to return to Section 2 (Research Design)]

Figure 4: 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)]

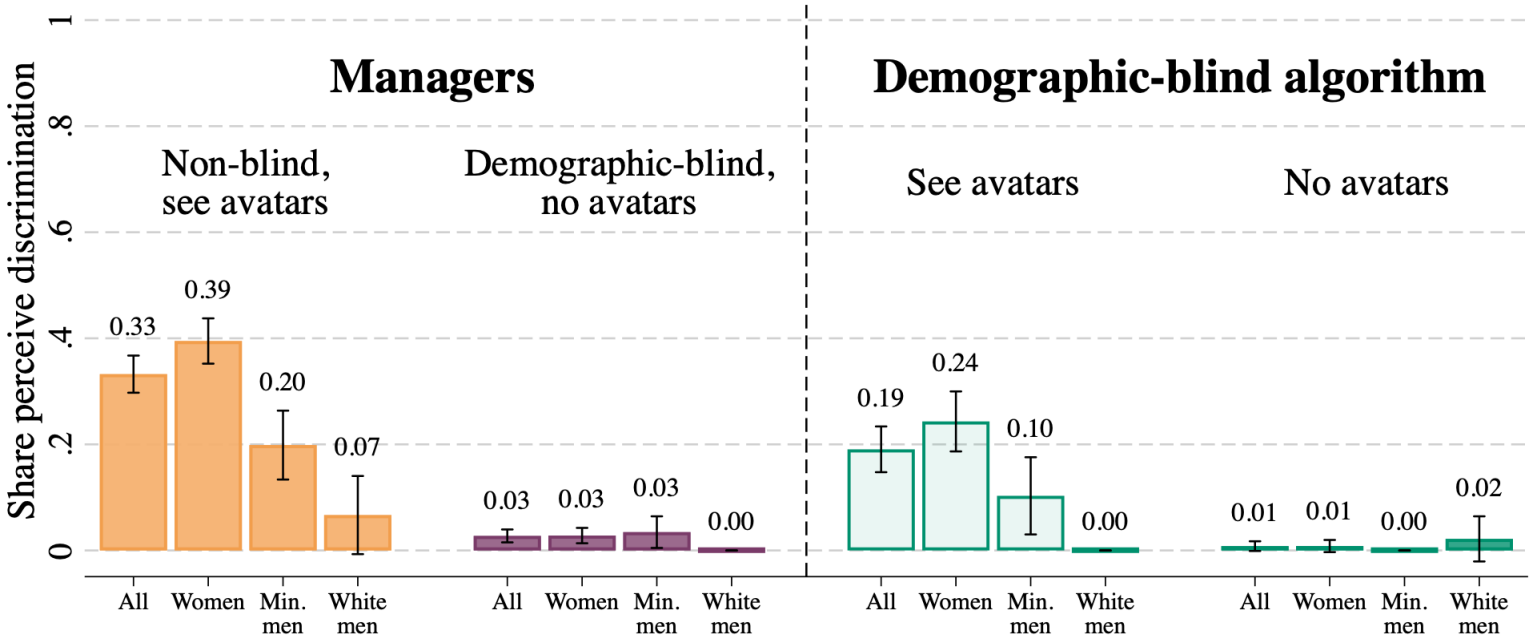
Figure 5: Design of hiring experiment



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.

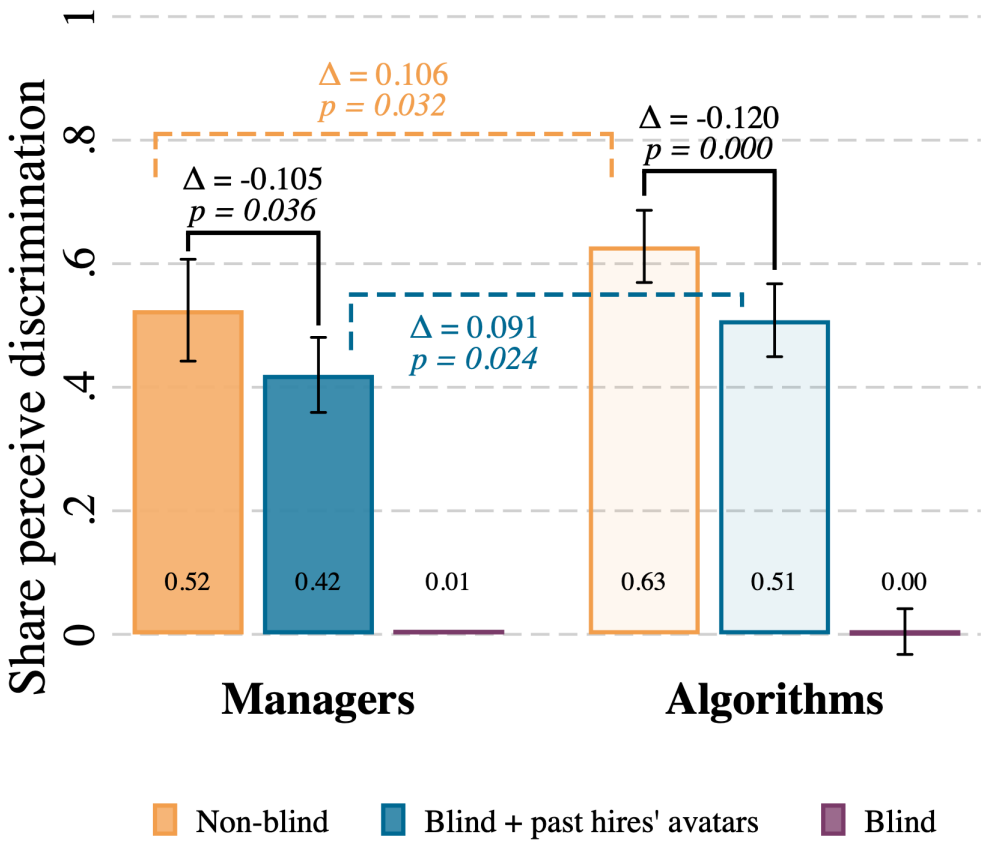
[Click to return to Section 1 (Introduction)]
[Click to return to Section 2 (Research Design)]

Figure 6: Perceived discrimination in the promotion experiment



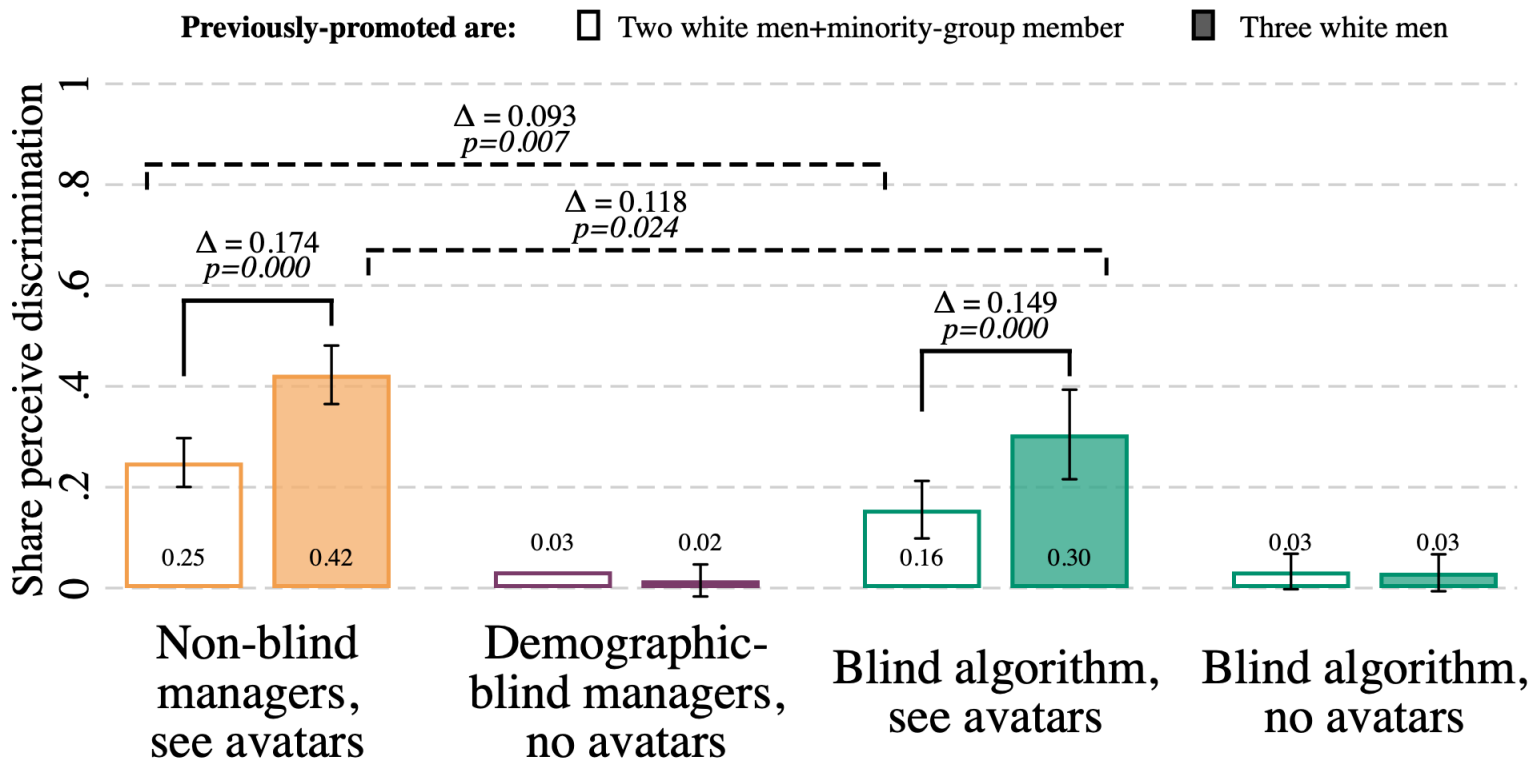
Note: This figure plots the share of workers in the experimental sample of the promotion experiment who perceive discrimination using the main measure (whether a worker mentions demographics in their response to the open-ended question, *"what needed to be different about your profile in order to be 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 in the left panel and in the two sub-arms of the algorithm treatment arm in the right panel. The figures restricts to workers who would not have been promoted under any promotion procedure. For breakdowns by race × gender, see Figure A37. 95 percent confidence intervals are indicated by the black bars. Treatment effects on this outcome with the full set of control variables are in Table 2. The corresponding effects on the secondary measures of perceived discrimination are in Appendix Tables A5 and A6.
[Click to return to Section 3.1 (Perceptions of Discrimination)]

Figure 7: Perceived discrimination in the hiring experiment



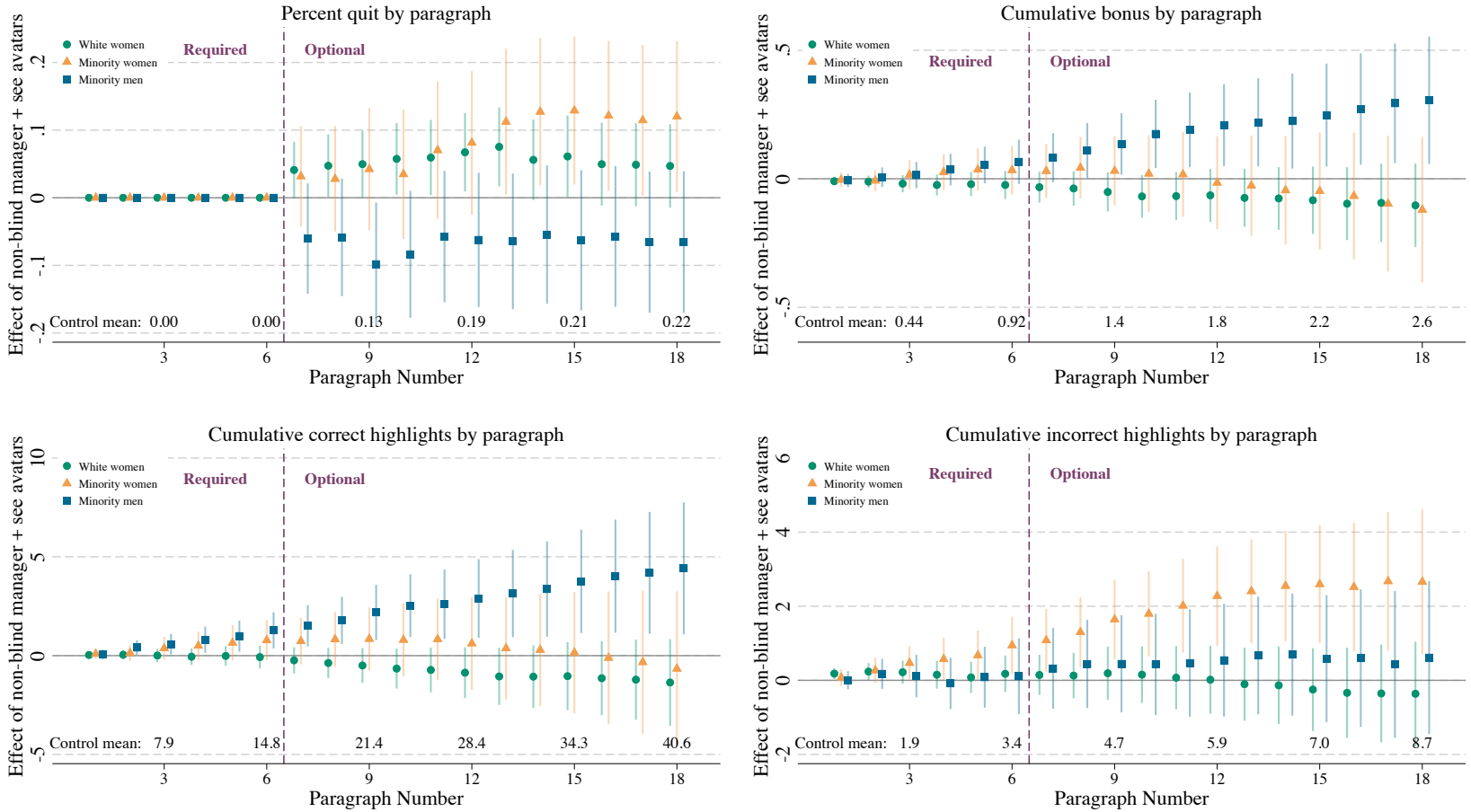
Note: This figure plots the share of workers perceiving discrimination in each treatment arm of the hiring experiment, using the main measure of perceived discrimination. The sample is restricted to workers who would not have been hired under any hiring procedure and workers who see three previous hires who are all white men (a random subsample within each treatment arm). For the analogous figure with the full sample (which does not account for differences across hiring procedures in the rate at which workers see three previous hires who are three white men versus seeing two white men and someone else), see Appendix Figure A5. 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 3.1 (Perceptions of Discrimination)]

Figure 8: Effects of seeing one previously-promoted minority-group worker in the promotion experiment



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. Note that the 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, if they had seen worker demographics. They do not see worker demographics. The figures restrict to workers who would not have been promoted under any procedure. 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 3.1 (Perceptions of Discrimination)]

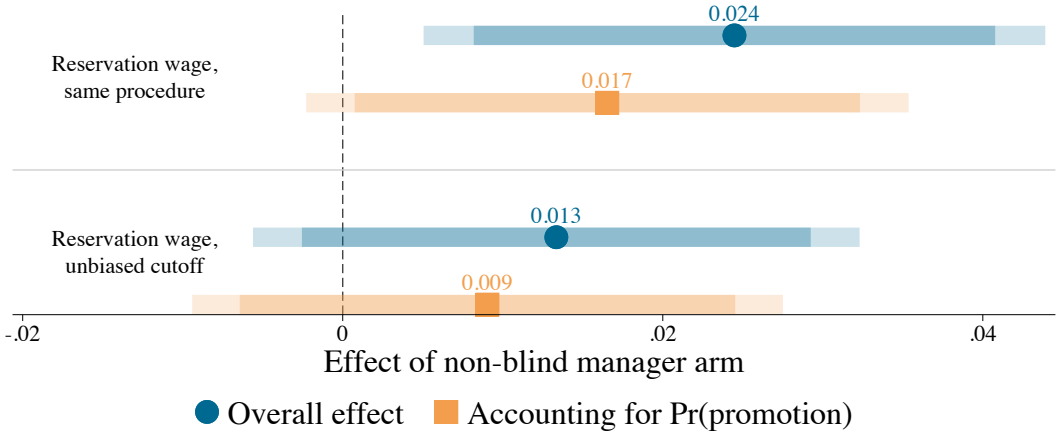
Figure 9: Effects on retention and overall earnings/performance in the promotion experiment by gender × race



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 whether workers have quit, how much they have earned, how many mistakes they have correctly identified, and how many non-mistakes they have incorrectly highlighted, each cumulatively by the given paragraph on the x-axis. Workers are required to proofread the first six paragraphs, after which they can quit after any paragraph, so treatment effects absent selection are available in the first six paragraphs. Effects are estimated separately for white women, racial minority women, and racial minority men. 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. Appendix Tables A7 and A8 present the corresponding effects on retention and the effects on performance in the first six paragraphs in Panel A and pool white and racial minority women in Panel B, including p-values testing the equality of the effects across demographic groups.

[Click to return to Section 3.2 (Retention and Performance)]

Figure 10: 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 that are most predictive of future labor supply, and 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 figures 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 shown, which are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.3 (Future Labor Supply)]

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: Perceptions of discrimination in the promotion experiment

	All workers	Women	Minority men	White men
Panel A: Manager sample, promotion experiment				
Non-blind manager, see avatars	0.307***	0.372***	0.174***	0.075
	(0.020)	(0.025)	(0.039)	(0.046)
Control mean	0.027	0.028	0.034	0.000
N	1389	1004	291	94
Panel B: Algorithm sample, promotion experiment				
See avatars	0.170***	0.220***	0.085*	-0.003
	(0.025)	(0.033)	(0.047)	(0.046)
Control mean	0.008	0.008	0.000	0.022
N	691	464	151	76

Note: This table reports treatment effects on the main measure of perceived discrimination (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?"*) for various demographic groups in the promotion experiment. The estimates for women pool women of all races and the estimates for racial minority men pool Asian, Black, and Hispanic men. Appendix Figure A37 disaggregates by race × gender. In panel A, 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. In panel B, the sample is those evaluated by the demographic-blind algorithm and the presented coefficient is on the indicator for being in the group that sees the race and gender of previously-promoted workers. The control group is workers evaluated by the same algorithm who don't see the demographics of previously-promoted workers. Both samples restrict to workers who would not have been promoted under any procedure. 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 3.1 (Perceptions of Discrimination)]

Table 3: Effects on retention

	Num paragraphs	More than 6	Did all 18
	(1)	(2)	(3)
Panel A: Manager sample, promotion experiment			
Non-blind manager,	-0.480*	-0.018	-0.045*
see avatars	(0.248)	(0.017)	(0.025)
N	1387	1387	1387
Control mean	15.916	0.915	0.776
Panel B: Algorithm sample, promotion experiment			
See avatars	0.139	0.029	0.005
	(0.396)	(0.029)	(0.039)
N	691	691	691
Control mean	15.456	0.881	0.730
p-value, manager = algorithm	*0.174*	*0.159*	*0.275*

Note: This table reports treatment effects in the promotion experiment on retention. 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. Samples and specifications are the same as in the first column of 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 3.2 (Retention and Performance)]

Table 4: Effects on effort and performance in the first six paragraphs

	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					
Non-blind manager,	-0.002	1.105	0.289	0.357*	-0.001
see avatars	(0.055)	(1.703)	(0.221)	(0.215)	(0.021)
N	1389	1389	1389	1389	1389
Control mean	4.957	36.804	14.817	3.422	0.919
Panel B: Algorithm sample, promotion experiment					
See avatars	0.061	-1.804	-0.703*	0.225	-0.059*
	(0.084)	(2.683)	(0.382)	(0.282)	(0.036)
N	691	691	691	691	691
Control mean	4.817	37.941	15.016	3.679	0.935
p-value, manager = algorithm	*0.521*	*0.348*	*0.021*	*0.703*	*0.149*

Note: This table reports treatment effects in the promotion experiment on effort and performance. The outcomes in the first and second columns are the total number of minutes spent on the first six required paragraphs and the total number of clicks on those paragraphs, respectively. The outcomes in columns 3, 4, and 5 are measures of performance: the number of correct highlights, number of incorrect highlights, and bonus earned in dollars in the first six required 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 3.2 (Retention and Performance)]

Table 5: Effects on future labor supply in the promotion experiment

	Res. wage (same)	WTP cutoff
	(1)	(2)
Panel A: Manager sample, promotion experiment		
Non-blind manager,	0.024**	0.010**
see avatars	(0.010)	(0.005)
N	1338	1325
Control mean	0.265	-0.014
Panel B: Algorithm sample, promotion experiment		
See avatars	0.011	0.001
	(0.015)	(0.006)
N	672	660
Control mean	0.282	-0.008
p-value, manager = algorithm	*0.454*	*0.234*

Note: This table reports treatment effects 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, i.e. it is the willingness to pay for the cutoff procedure. Both reservation wages are elicited over the easier and harder job wages simultaneously, with the harder job wage equal to 2 times the easier job wage, and in in this table they are constructed assuming that each worker thought there was a 50 percent chance of being assigned to the harder job (as was the approximate average in the control groups). 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 (Future Labor Supply)]

Table 6: Effects on sharing with and avoidance of manager in the promotion experiment

	Cooperative task RW	WTP new manager	Shared zero	Amt bonus shared
	(1)	(2)	(3)	(4)
Non-blind manager, see avatars	0.010	0.043**	0.049*	-0.302
	(0.014)	(0.020)	(0.028)	(0.231)
N	1349	1030	1383	1383
Control mean	0.341	0.223	0.522	2.857

Note: 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. The specifications and sample are the same as in Panel A of 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 (Future Labor Supply)]

Table 7: Effects on beliefs about promotion

	Pr(Promoted|Same)	Pr(Promoted|Cutoff)	Diff Pr(Promoted)
	(1)	(2)	(3)
Panel A: Manager sample, promotion experiment			
Non-blind manager, see avatars	-3.079***	-1.650	-1.430**
	(1.124)	(1.172)	(0.699)
N	1385	1385	1385
Control mean	47.219	48.491	-1.272
Panel B: Algorithm sample, promotion experiment			
See avatars	-0.463	-1.487	1.024
	(1.674)	(1.819)	(0.956)
N	689	689	689
Control mean	45.908	48.201	-2.293
p-value, manager = algorithm	*0.184*	*0.939*	*0.034*

Note: This table reports treatment effects 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 (Future Labor Supply)]

Table 8: 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 A11 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 7 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.
[Click to return to Section 3.3 (Future Labor Supply)]

Online Appendix: Perceived discrimination at work

Hannah Ruebeck

November 2023

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 Deviations from the pre-analysis plan

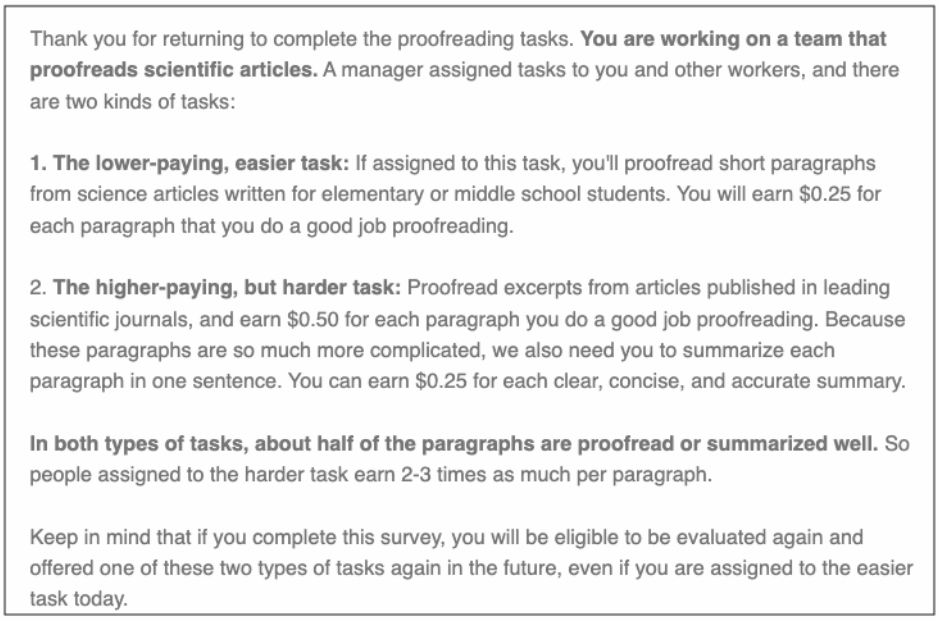
H Coding the text data

I Implicit measures of perceived discrimination

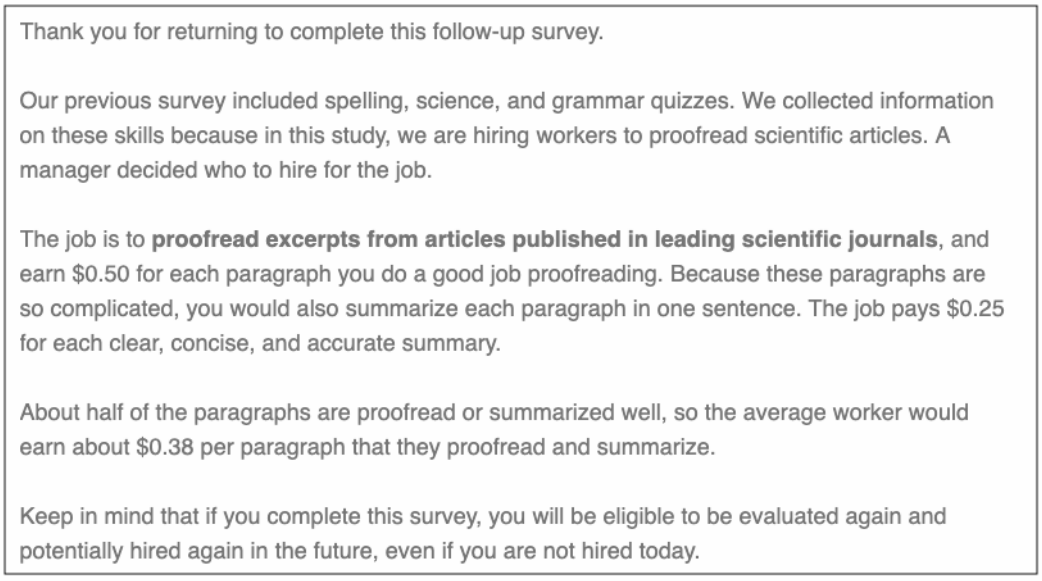
A Supplementary figures and tables

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

Panel A: Promotion experiment

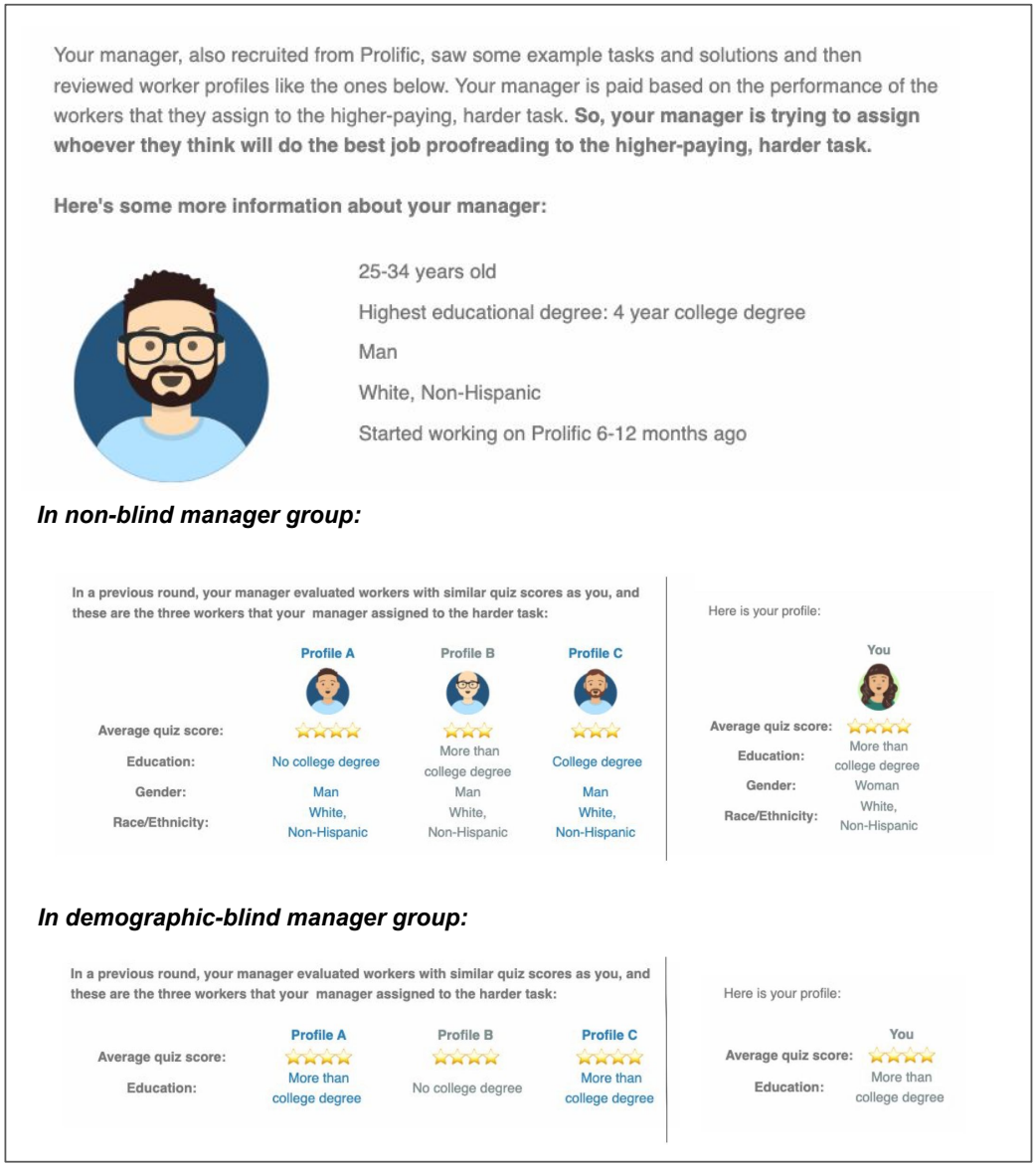


Panel B: Hiring experiment



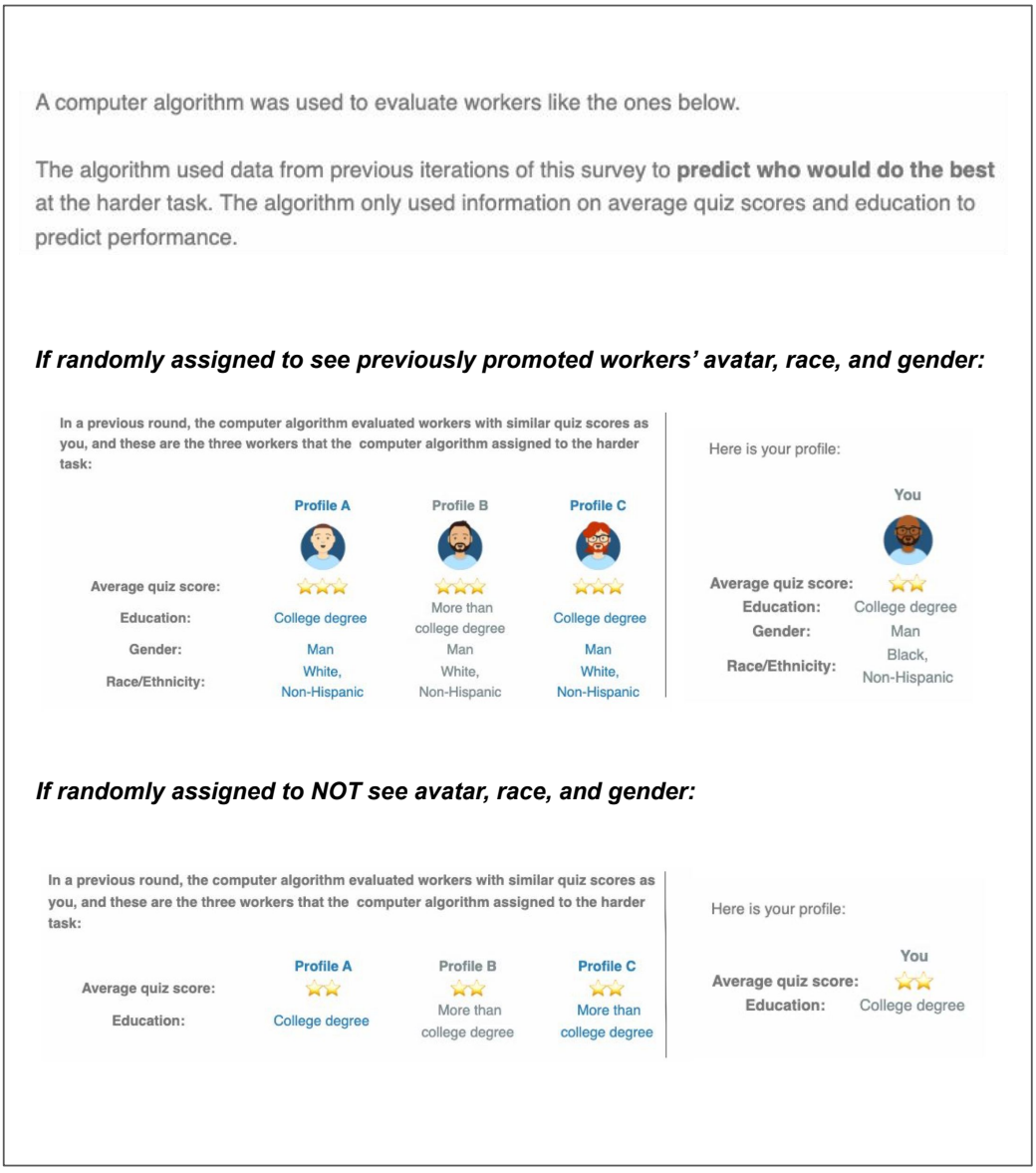
Note: This figure shows the information that workers received about the easier, lower-paying ("non-promotion") job and the harder, higher-paying ("promotion") job in the promotion experiment (Panel A) or the job that they were not hired to do in the hiring experiment (Panel B).
[Click to return to Section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

Figure A2: Treatment variation in manager groups



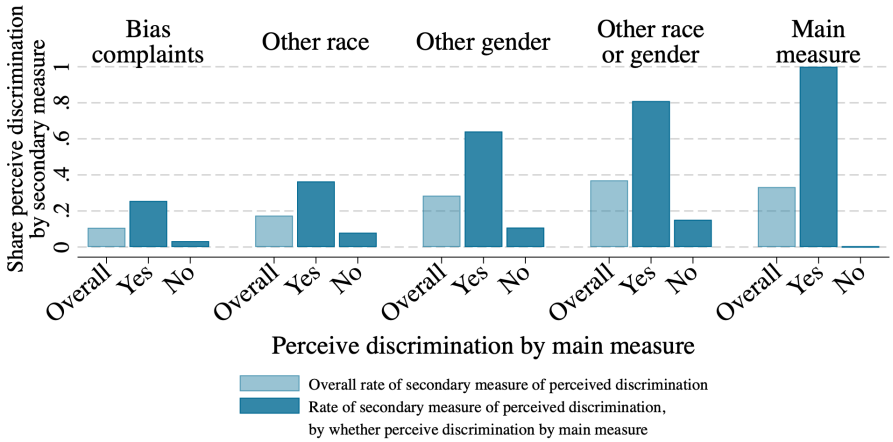
Note: This figure shows what workers saw about their manager, regardless of whether their manager saw avatars, race, and gender or not, and the information they saw about the workers their manager previously promoted in the non-blind manager group and demographic-blind manager group.
[Click to return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

Figure A3: Treatment variation in algorithm groups (promotion experiment)



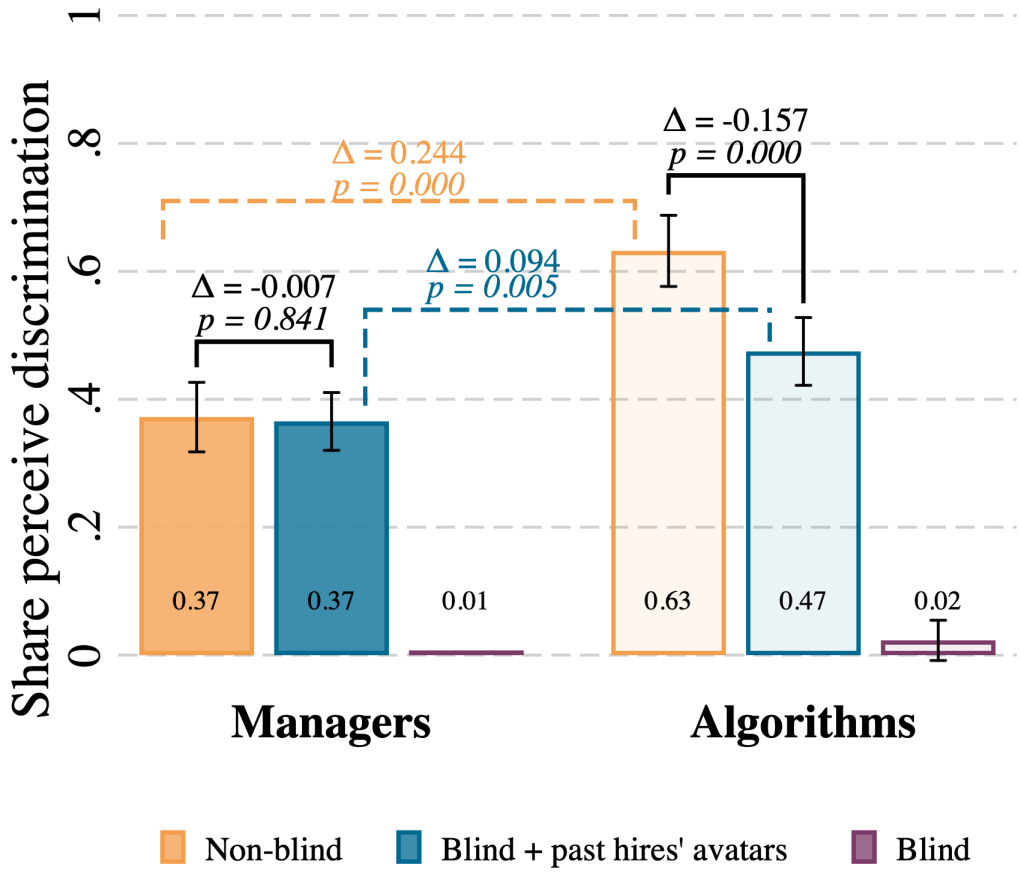
Note: This figure shows what workers saw about the algorithm and the information they saw about the workers their manager previously promoted depending on whether they were randomly assigned to see those workers' avatars, race, and gender or not.
[Click to return to Section 2 (Research Design)]

Figure A4: Correlation of perceived discrimination measures



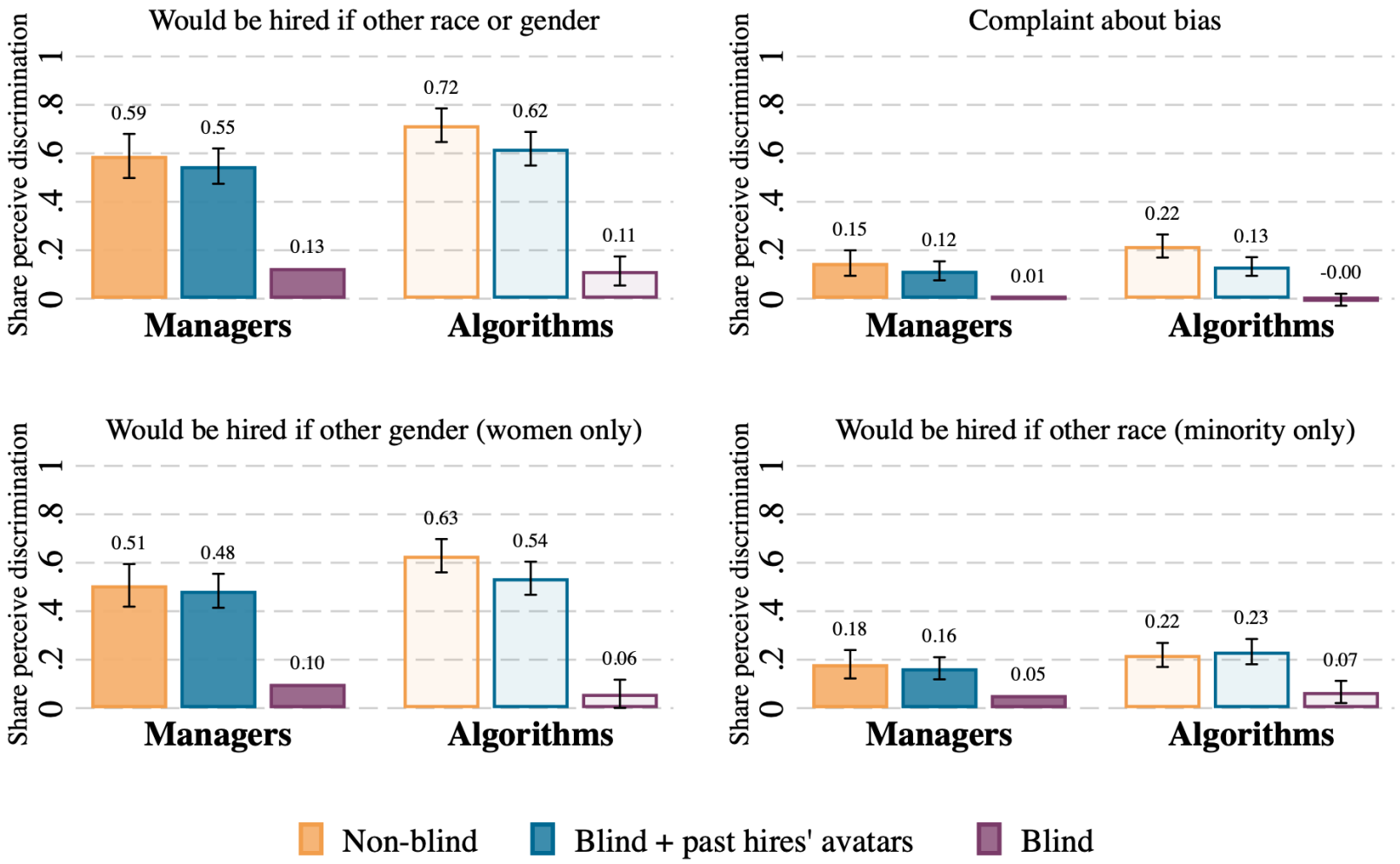
Note: This figure shows the correlation between the primary and secondary measures of perceived discrimination and the overall frequency of each in the non-blind manager arm of the promotion experiment. The primary measure is an indicator for whether a worker mentions demographics in their response to 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 (probably or definitely) have been promoted if they had a different race, an analogous indicator for gender, and an indicator that is the maximum of the two. The sample is those in the non-blind manager arm of the promotion experiment who would not have been promoted under any procedure, but the correlation is the same in other arms with positive rates of perceived discrimination. 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 to illustrate the correlations.
[Click to return to Section 2 (Research Design)]

Figure A5: Perceived discrimination in the hiring experiment (full sample)



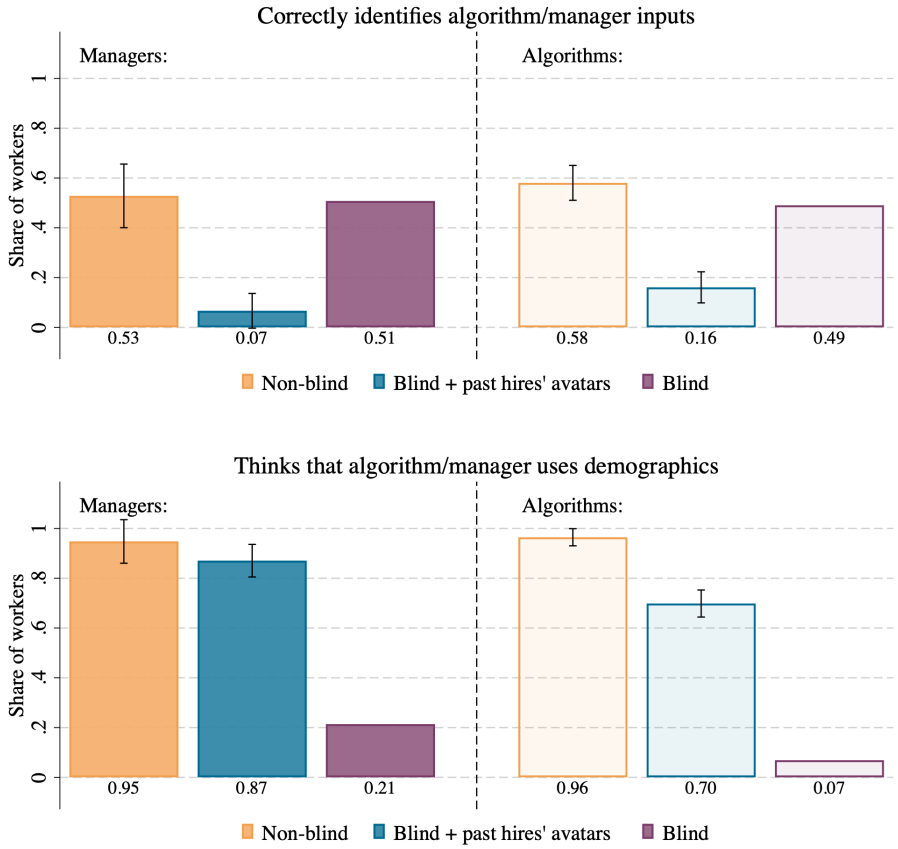
Note: This figure replicates Figure 7 but for all workers regardless of the demographic makeup of the previously-hired workers they saw, without accounting for differences across hiring procedures in the rate at which workers saw that all three previous hires were three white men versus seeing two white 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 3.1 (Perceived Discrimination)]

Figure A6: Perceived discrimination in the hiring sample, secondary measures of perceived discrimination



Note: This figure replicates Figure 7 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 A4. 95 percent confidence intervals (in black bars) are calculated with standard errors robust to heteroskedasticity.
[Click to return to Section 3.1 (Perceived Discrimination)]

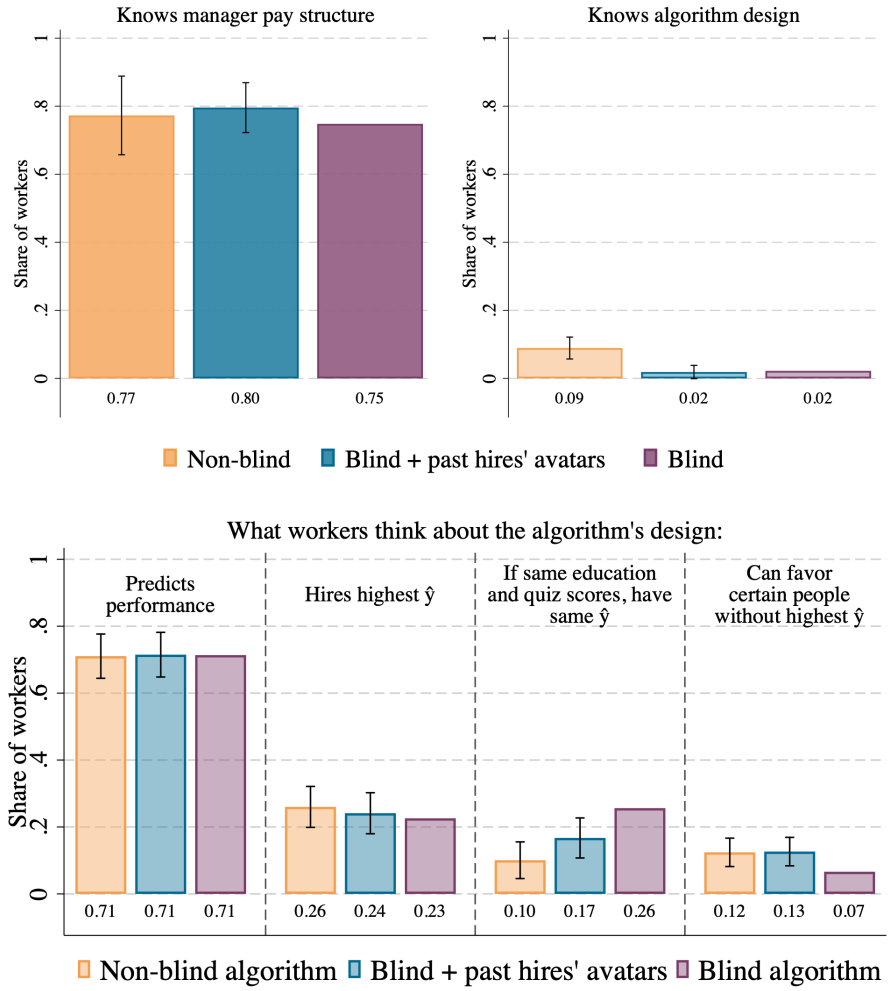
Figure A7: Comprehension of hiring procedure inputs



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.

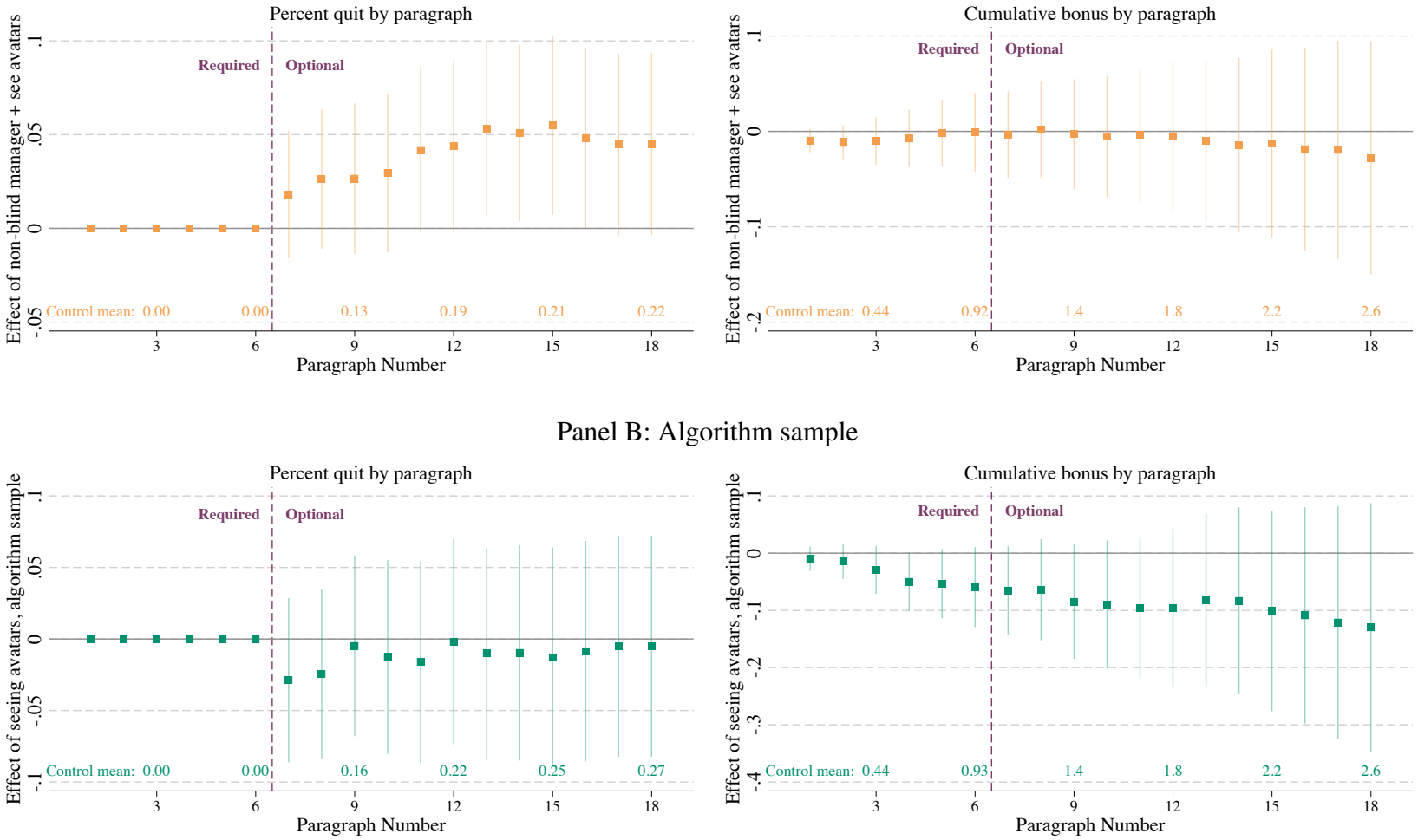
[Click to return to Section 3.1 (Perceived Discrimination)]

Figure A8: Comprehension of hiring procedure incentives (manager arms) and design (algorithm arms) in the hiring experiment



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 3.1 (Perceived Discrimination)]

Figure A9: 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.

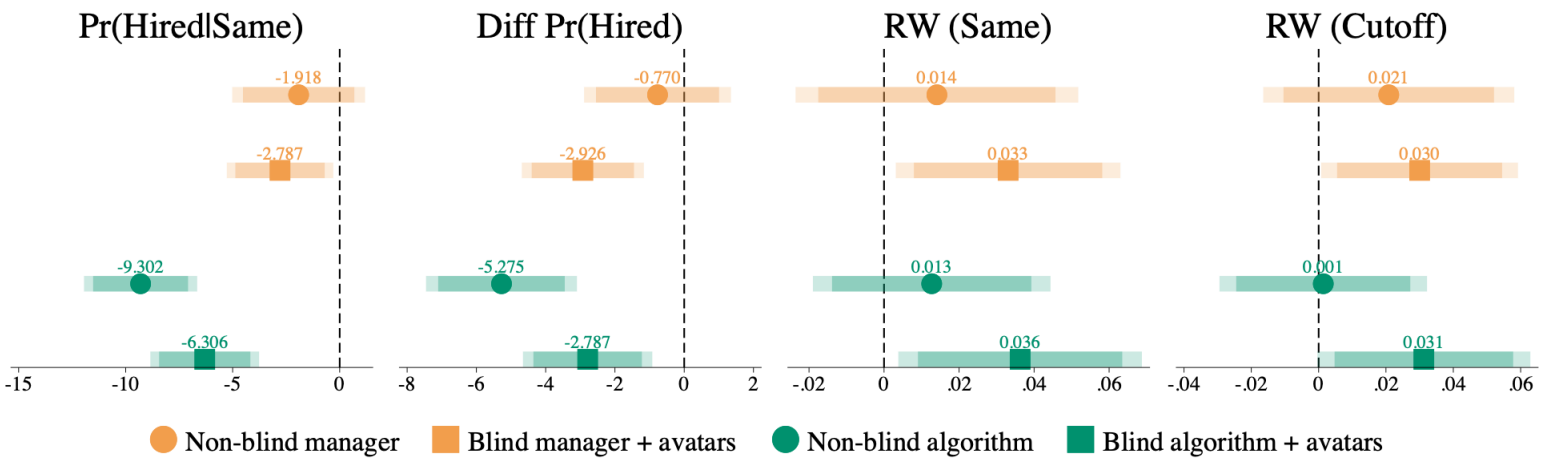
[Click to return to Section 3.2 (Retention and Performance)]

Figure A10: 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 3.2 (Retention and Performance)]

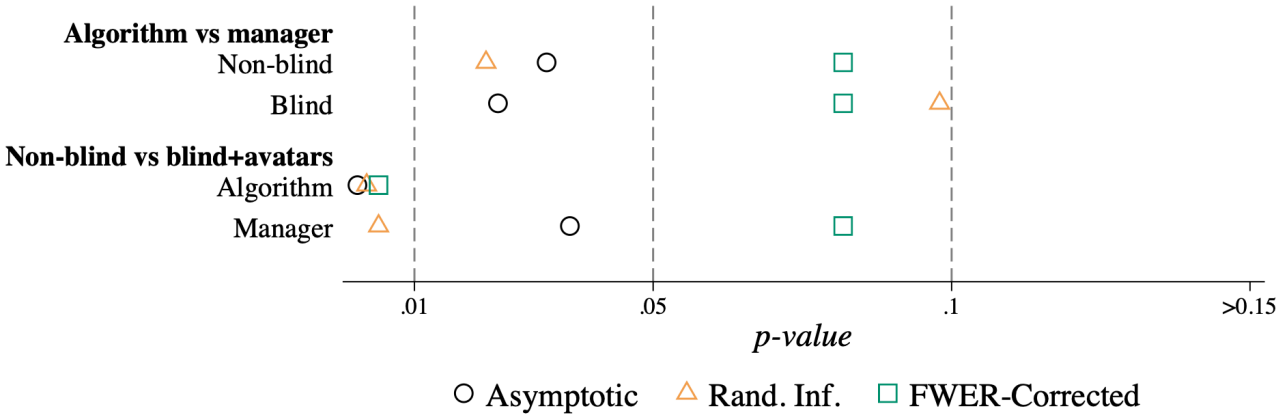
Figure A11: Effects on beliefs, reservation wages, and affect in the hiring experiment, separately by treatment arm



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 8—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.

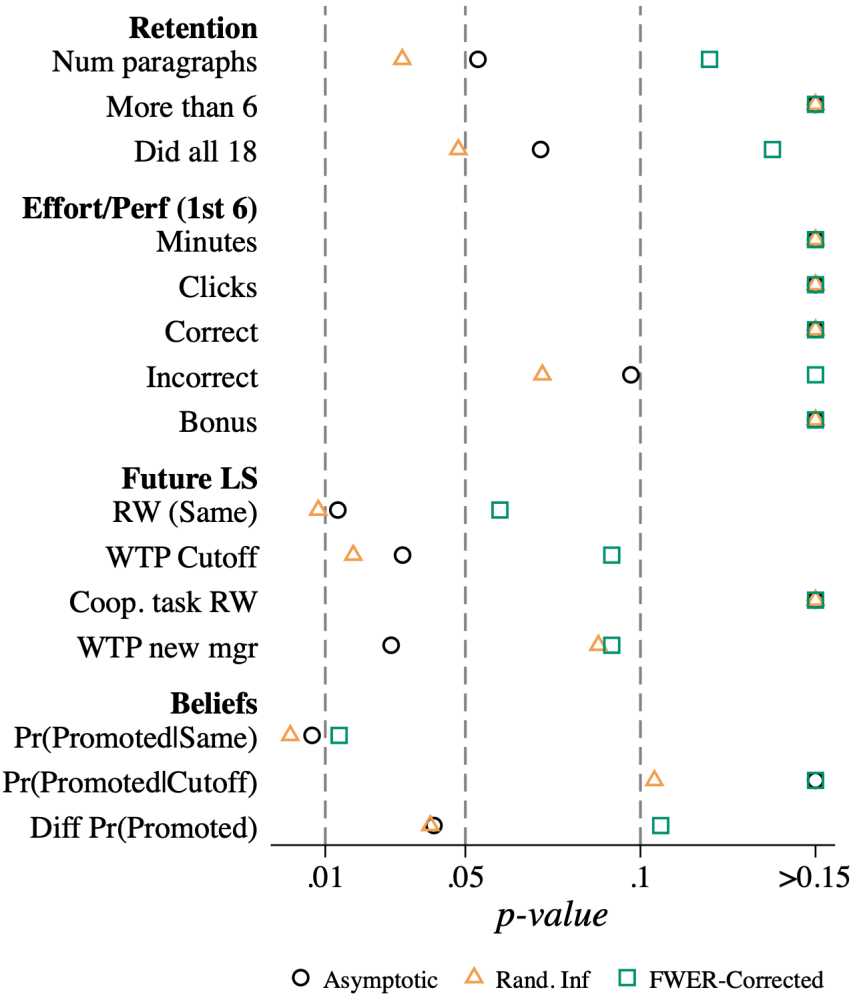
[Click to return to Section 3.3 (Future Labor Supply)]

Figure A12: Randomization inference and multiple hypothesis testing (hiring experiment)



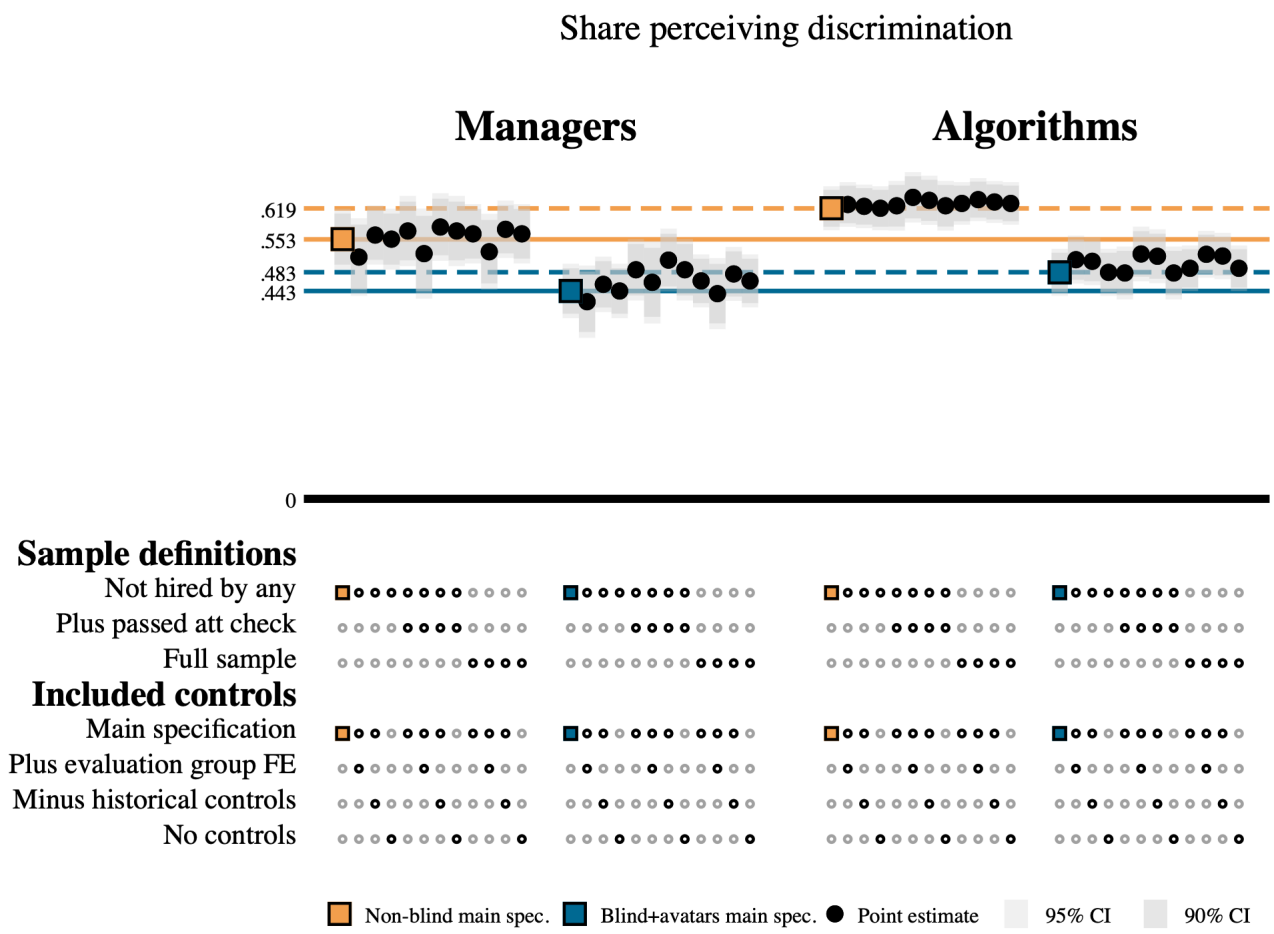
Note: This figure plots the *p*-values for the differences in perceived discrimination between treatment arms in Figure 7, with the *p*-values calculated asymptotically (as in the main analysis, they are robust to heteroskedasticity), with randomization inference, and corrected for multiple hypothesis testing using the Romano-Wolf correction.
[Click to return to Section 4 (Robustness)]

Figure A13: Randomization inference and multiple hypothesis testing (promotion experiment)



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 3, 4, 5, and 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.
[Click to return to Section 4 (Robustness)]

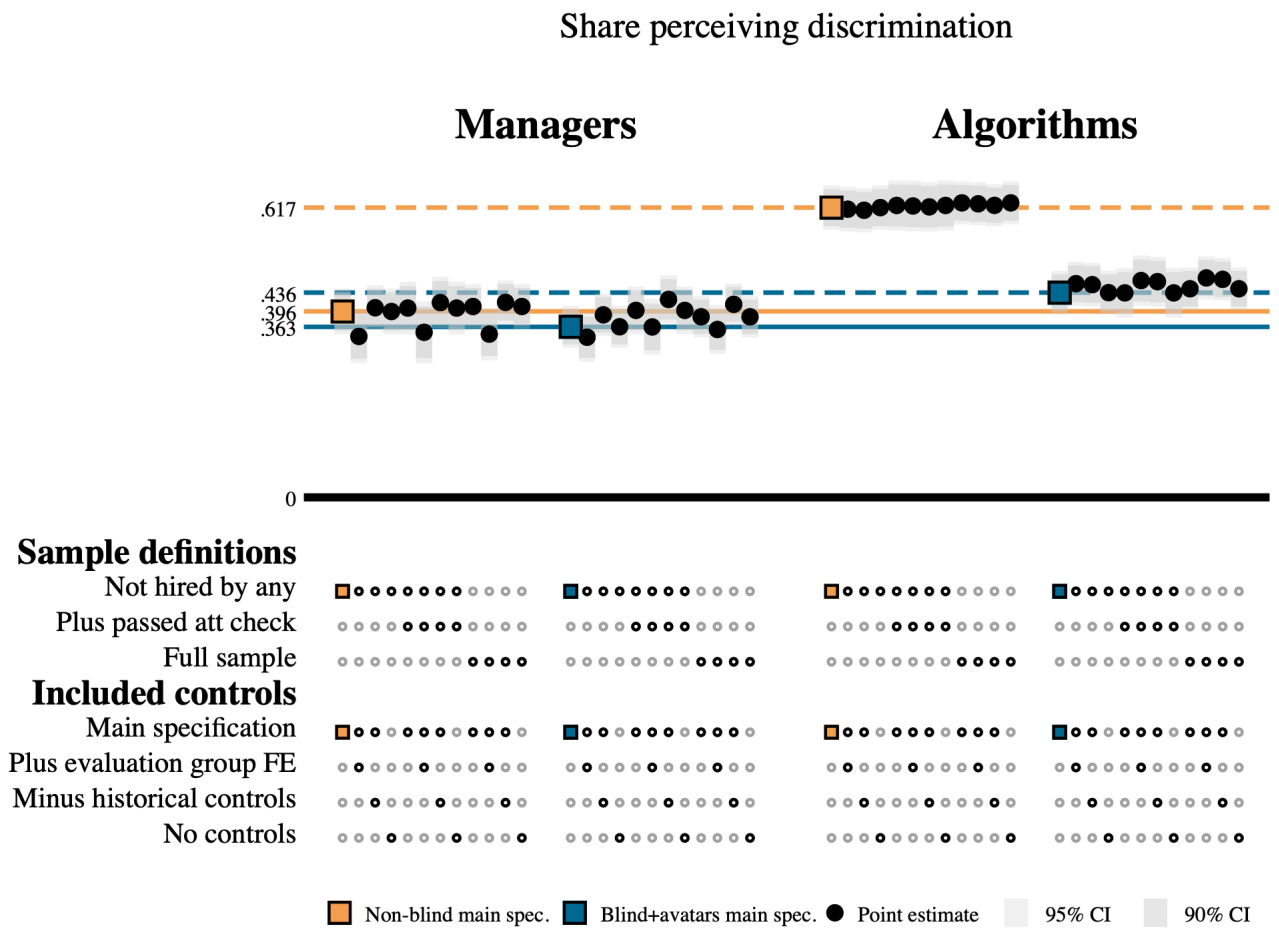
Figure A14: Alternative specifications: Hiring experiment (if previous hires are all white men)



Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the hiring experiment. It parallels Figure 7, 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.

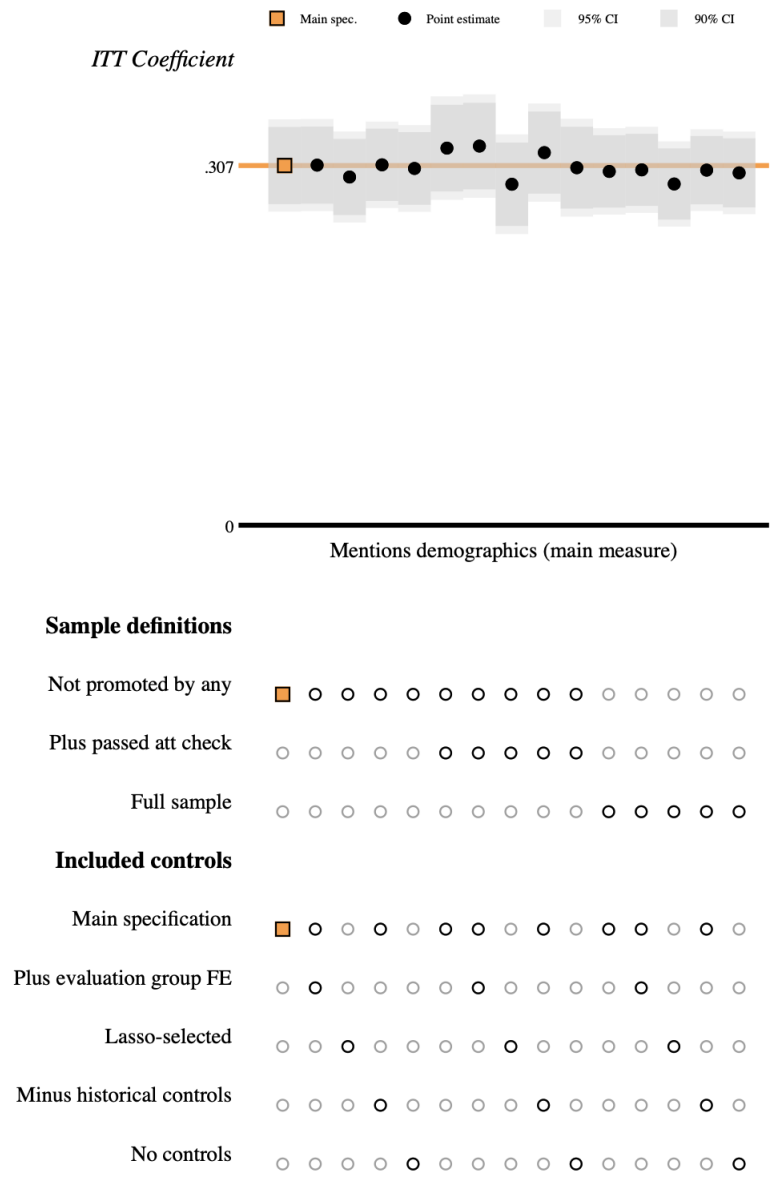
[Click to return to Section 4 (Robustness)]

Figure A15: Alternative specifications: Hiring experiment (regardless of demographic makeup of previous hires)



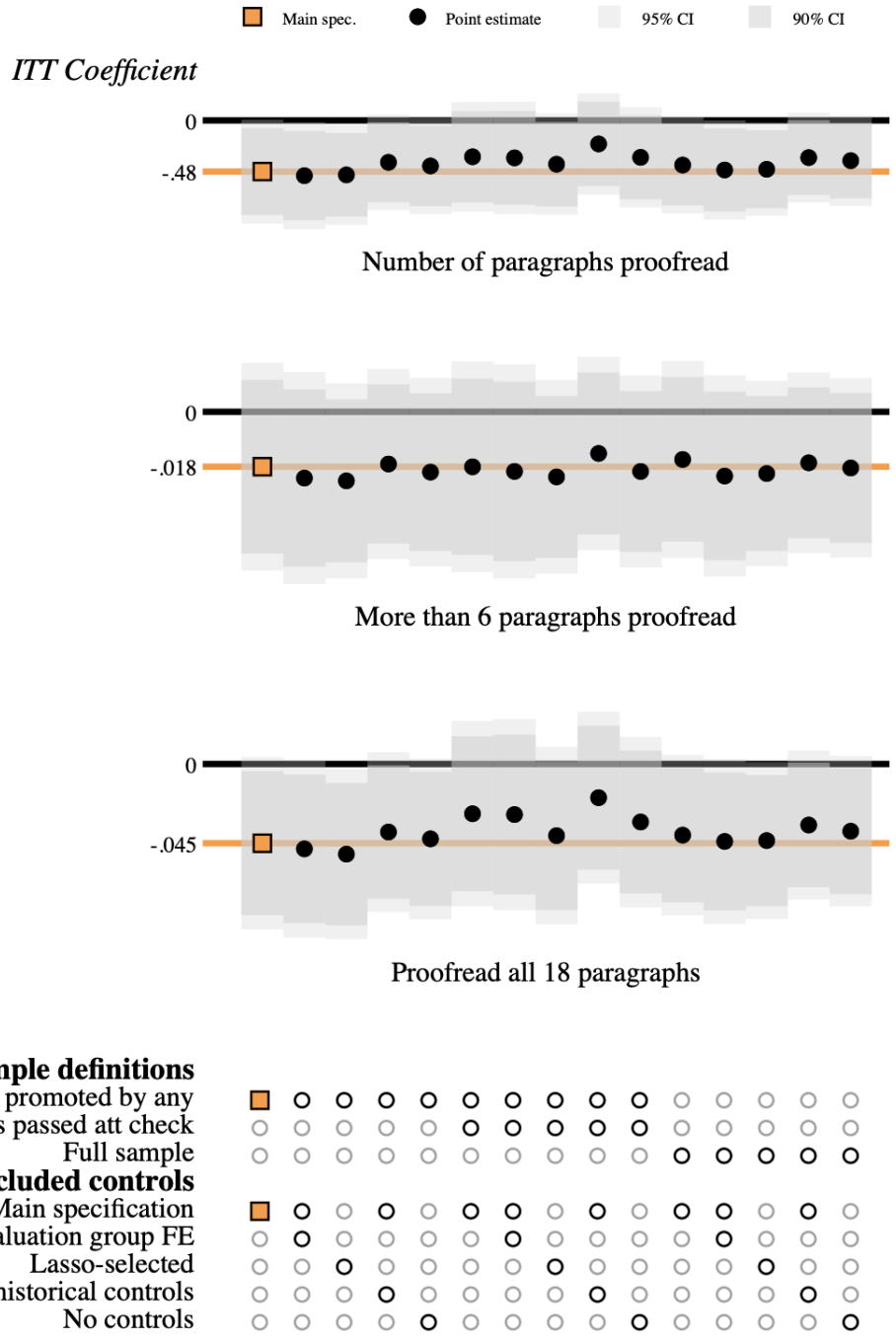
Note: This figure plots the estimated share of workers perceiving discrimination by the main measure in the hiring experiment, paralleling Appendix Figure A5 in using the full sample (not accounting for differences in the share of previous hires who are white men). Otherwise, it is analogous to Appendix Figure A14.
[Click to return to Section 4 (Robustness)]

Figure A16: Alternative specifications: Effect of being in non-blind manager arm on perceived discrimination (promotion experiment)



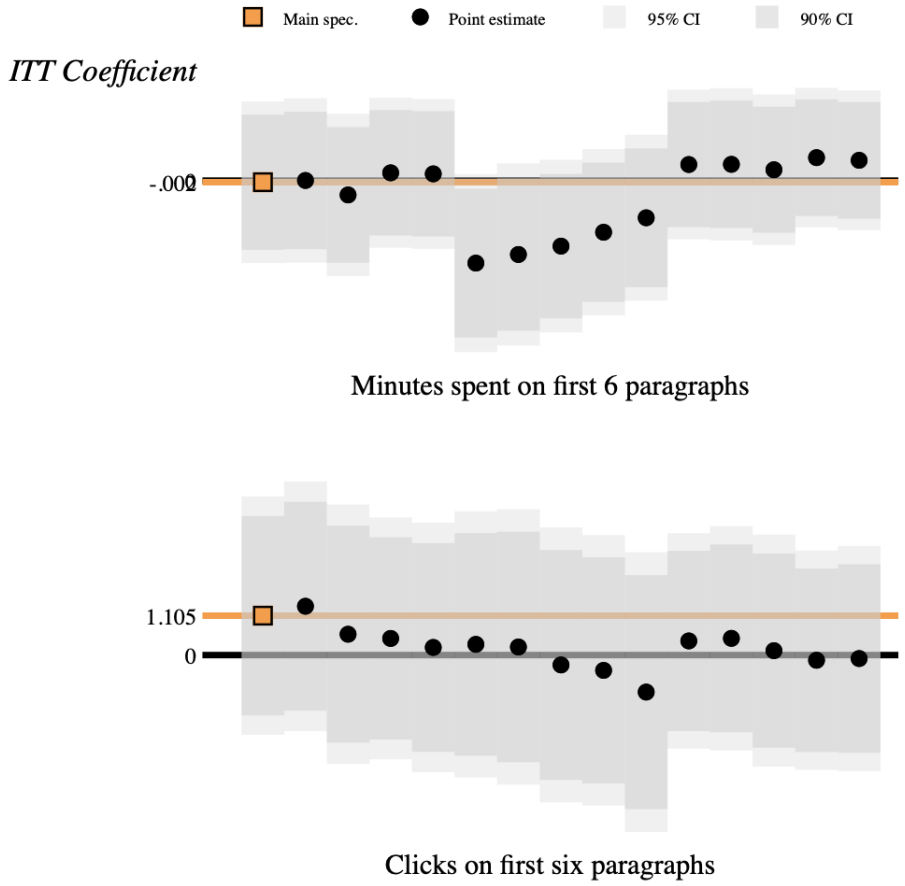
Note: This figure plots the treatment effect of being in the non-blind manager group on the main measure of perceived discrimination in the promotion experiment, relative to the demographic-blind manager group. The figure iterates over sample restrictions and choices of control variables, starting with 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.
[Click to return to Section 4 (Robustness)]

Figure A17: Alternative specifications: Effect of being in non-blind manager arm on retention



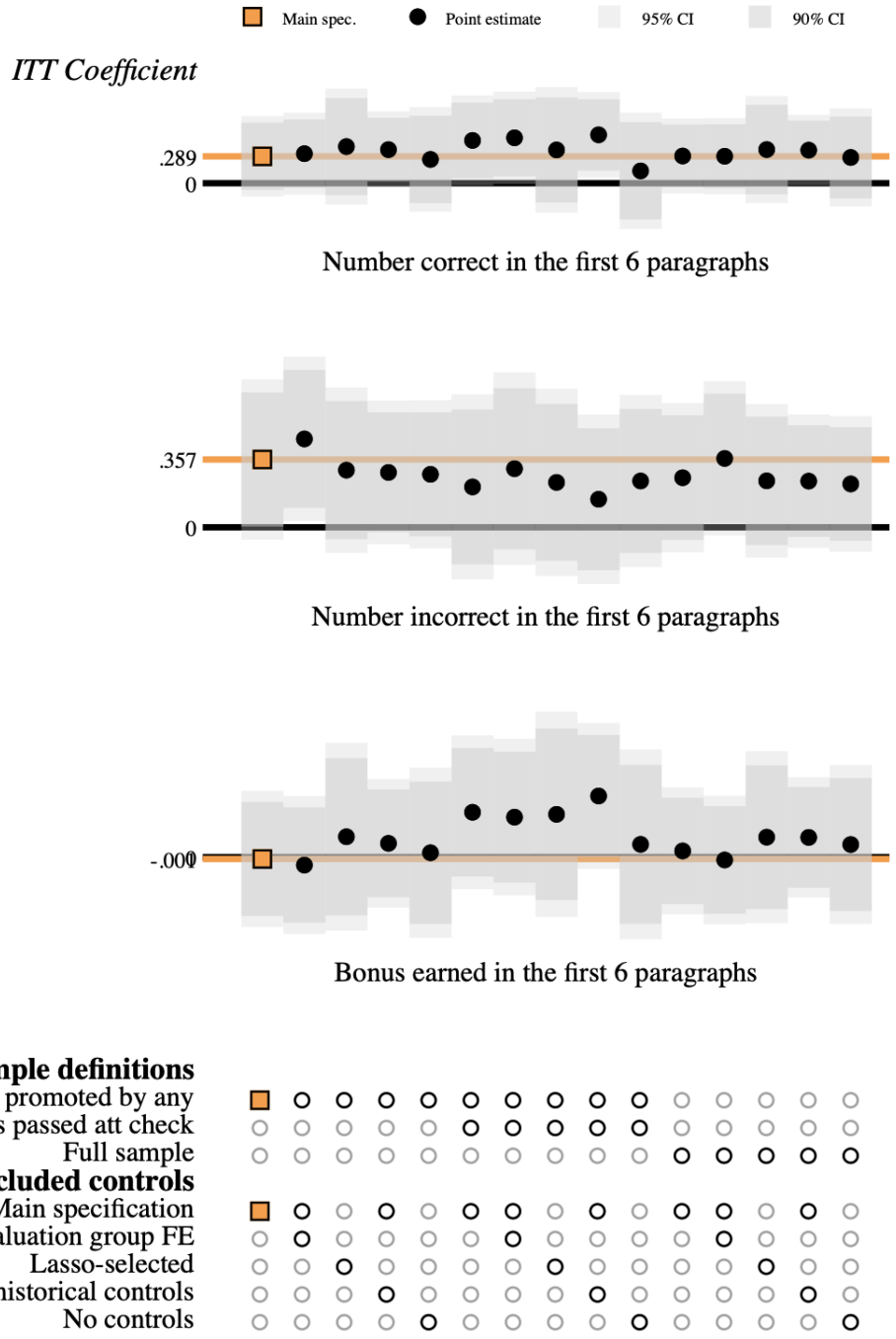
Note: This figure plots the treatment effect of being in the non-blind manager group on retention in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 3. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A18: Alternative specifications: Effect of being in non-blind manager arm on effort



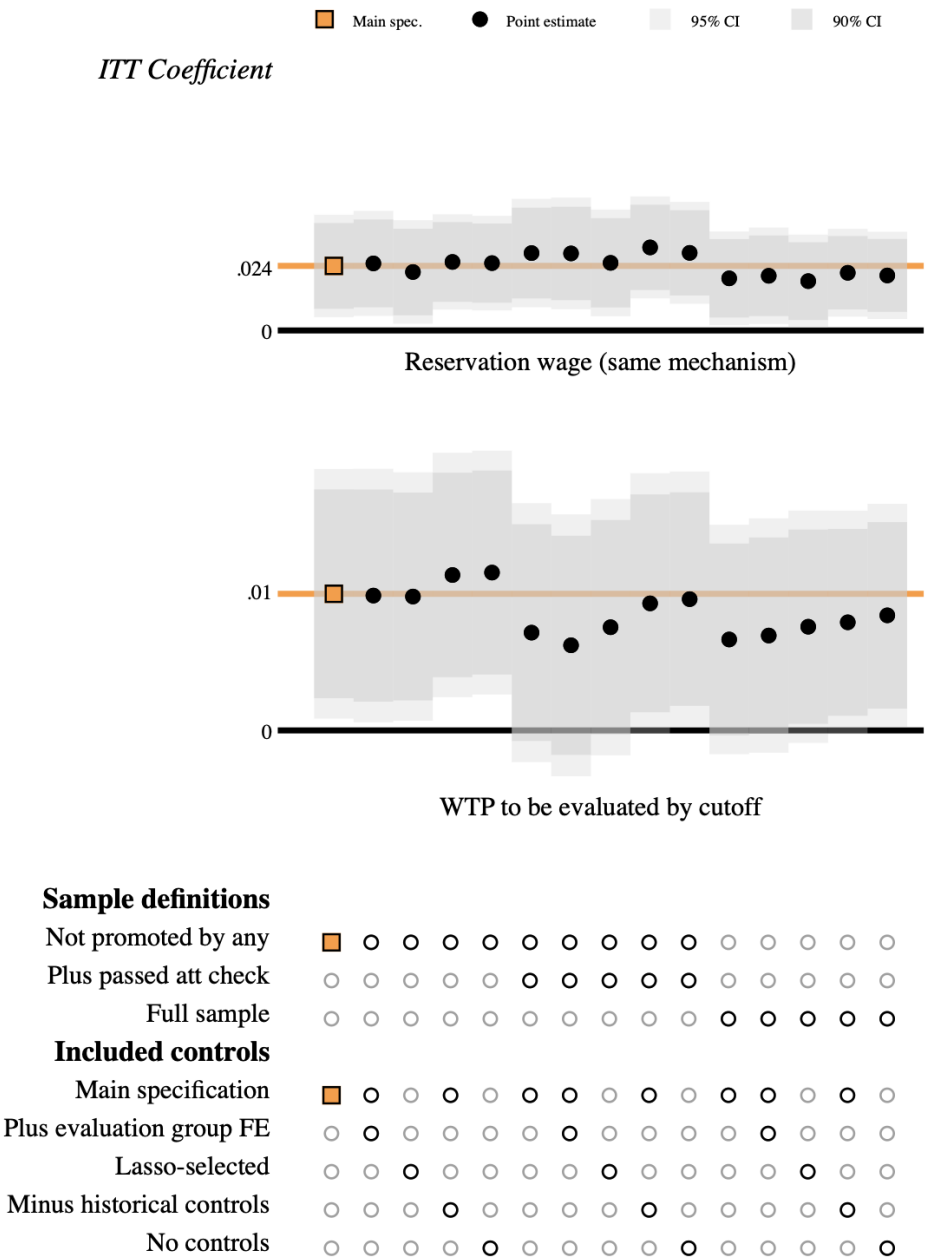
Note: This figure plots the treatment effect of being in the non-blind manager group on effort in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in columns 1 and 2 of Table 4. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A19: Alternative specifications: Effect of being in non-blind manager arm on performance



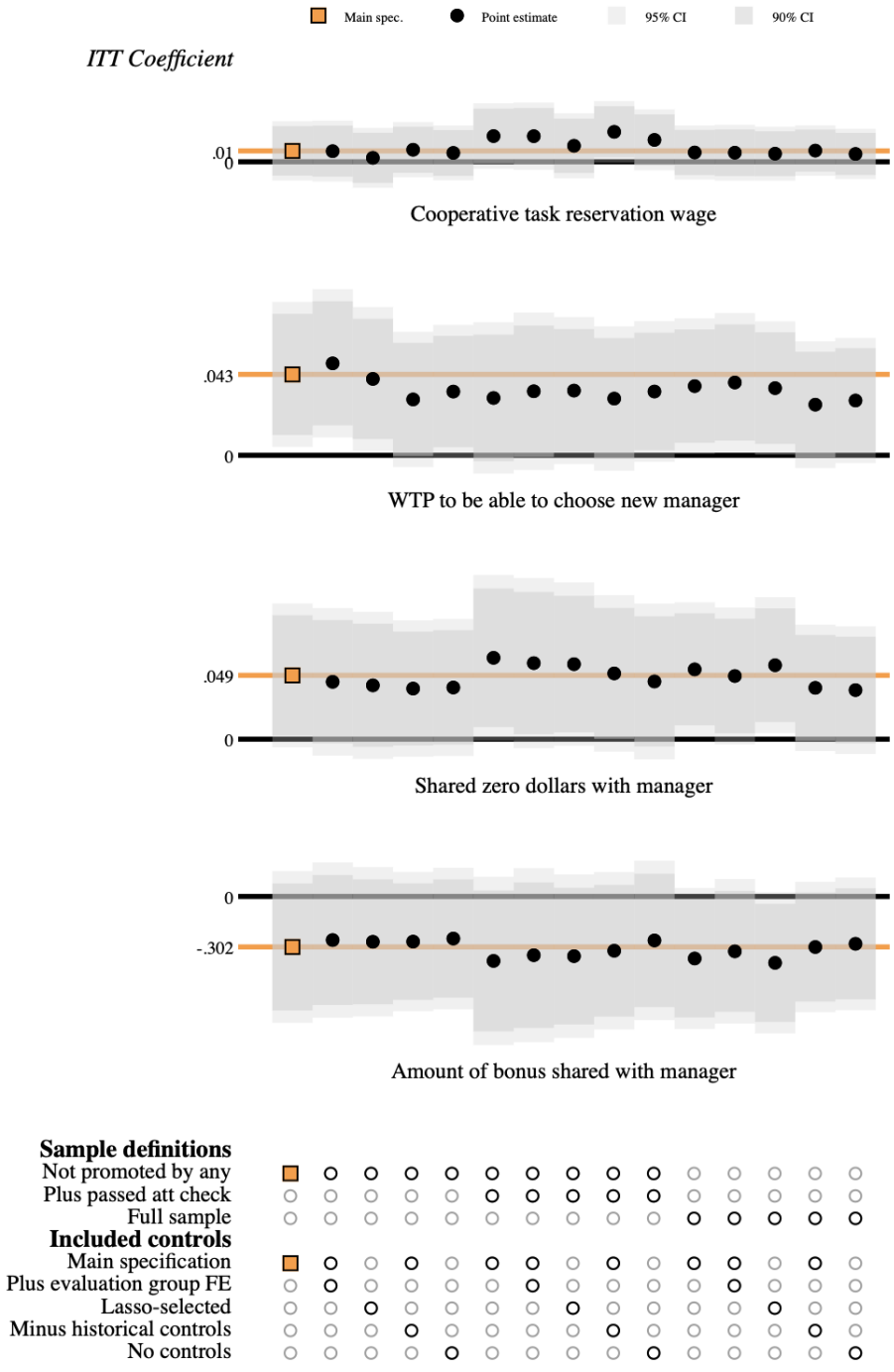
Note: This figure plots the treatment effect of being in the non-blind manager group on performance in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in columns 3-5 of Table 4. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A20: Alternative specifications: Effect of being in non-blind manager arm on future labor supply for the same job



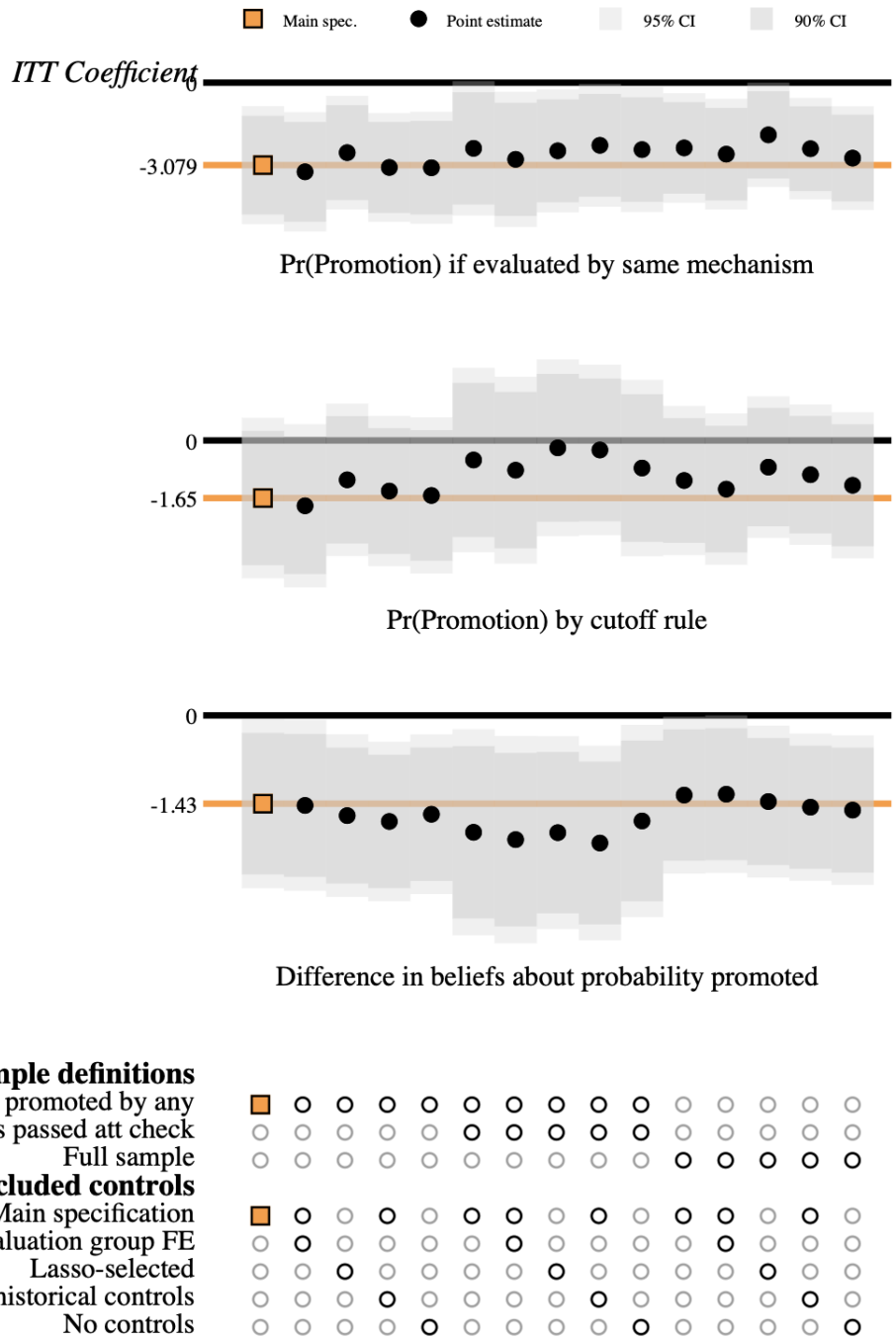
Note: This figure plots the treatment effect of being in the non-blind manager group on reservation wages in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 5. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A21: Alternative specifications: Effect of being in non-blind manager arm on future labor supply (cooperative task) and generosity



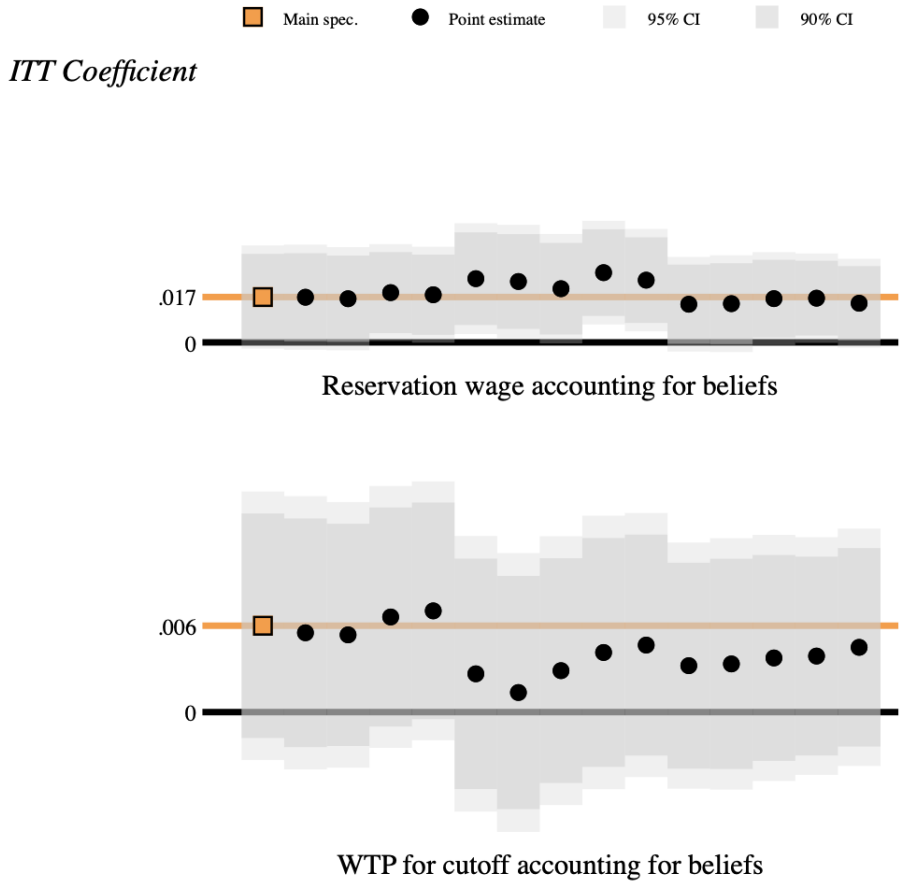
Note: This figure plots the treatment effect of being in the non-blind manager group on willingness to interact and share with managers in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 6. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A22: Alternative specifications: Effect of being in non-blind manager arm on beliefs about promotion



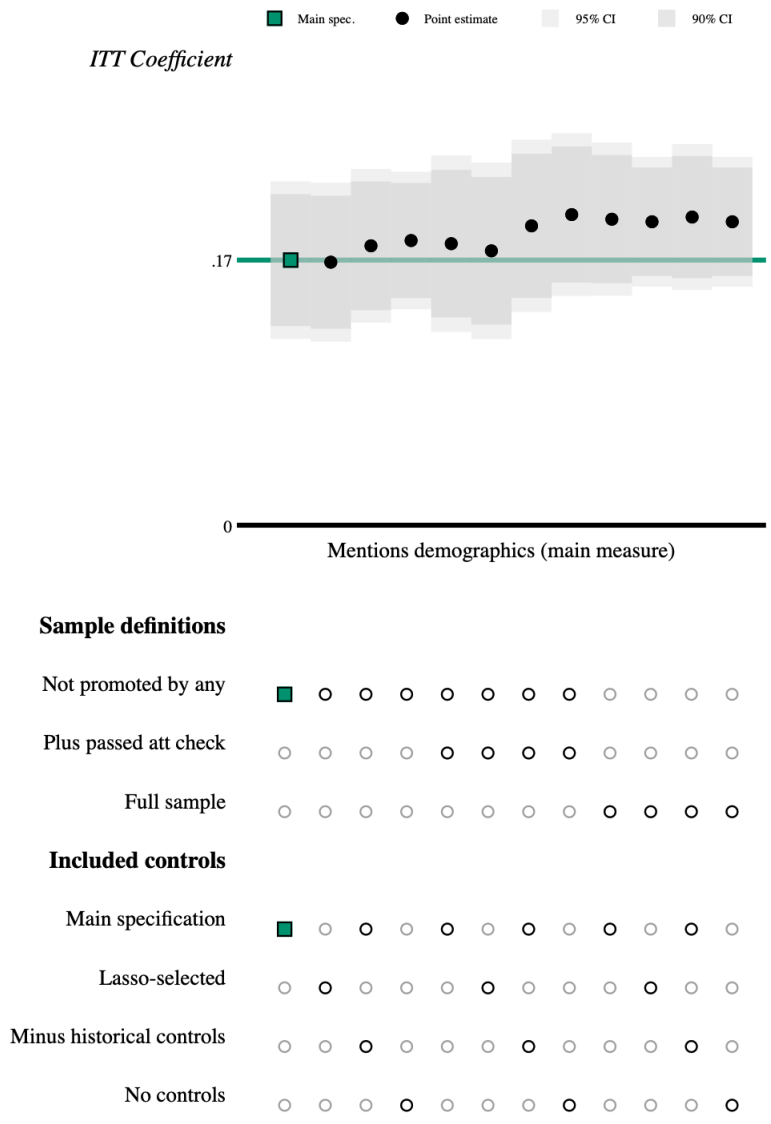
Note: This figure plots the treatment effect of being in the non-blind manager group on beliefs about promotion in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in Table 7. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A23: Alternative specifications: Effect of being in non-blind manager arm on future labor supply, accounting for beliefs



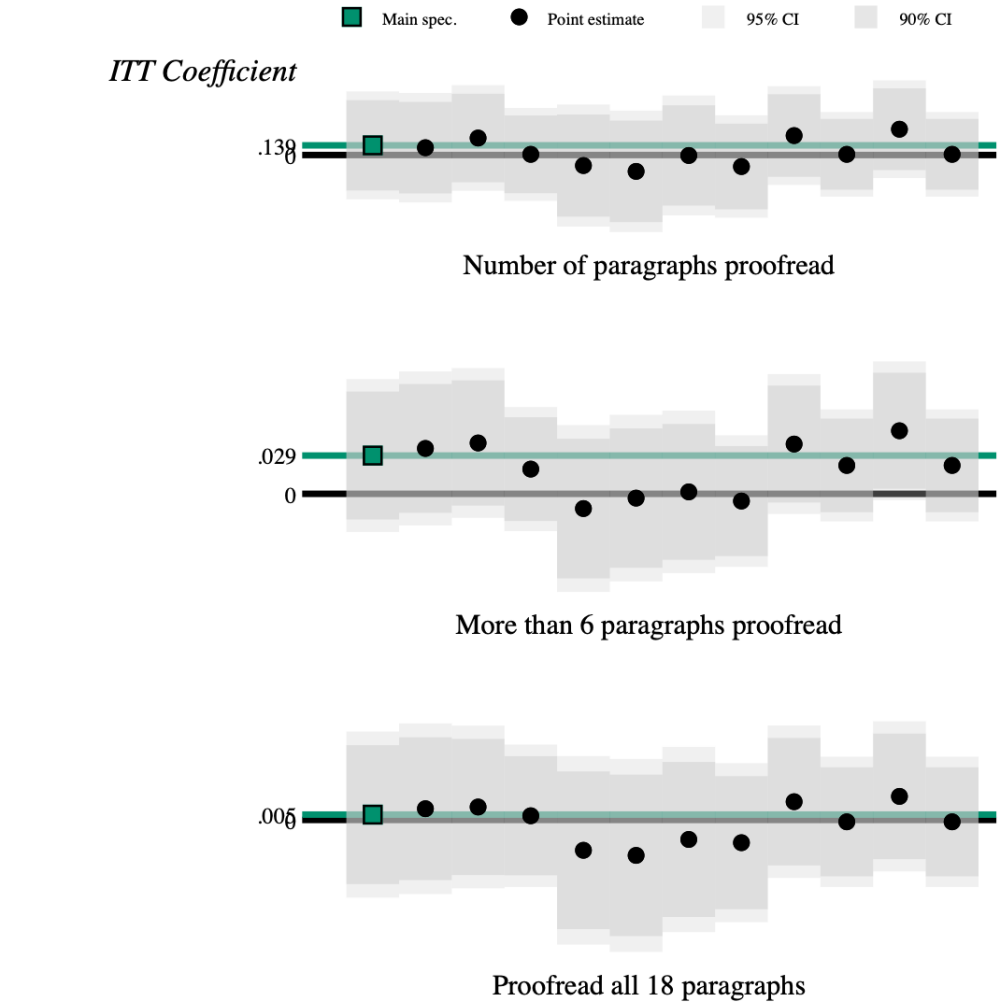
Note: This figure plots the treatment effect of being in the non-blind manager group on reservation wages after accounting for the effects on beliefs about promotion in the promotion experiment, relative to the demographic-blind manager group. The outcomes and main specification are those in the second specification (in blue) in Figure 10. Otherwise, the figure is analogous to Appendix Figure A16.
[Click to return to Section 4 (Robustness)]

Figure A24: Alternative specifications: Effects of seeing avatars on perceived algorithmic discrimination (promotion experiment)



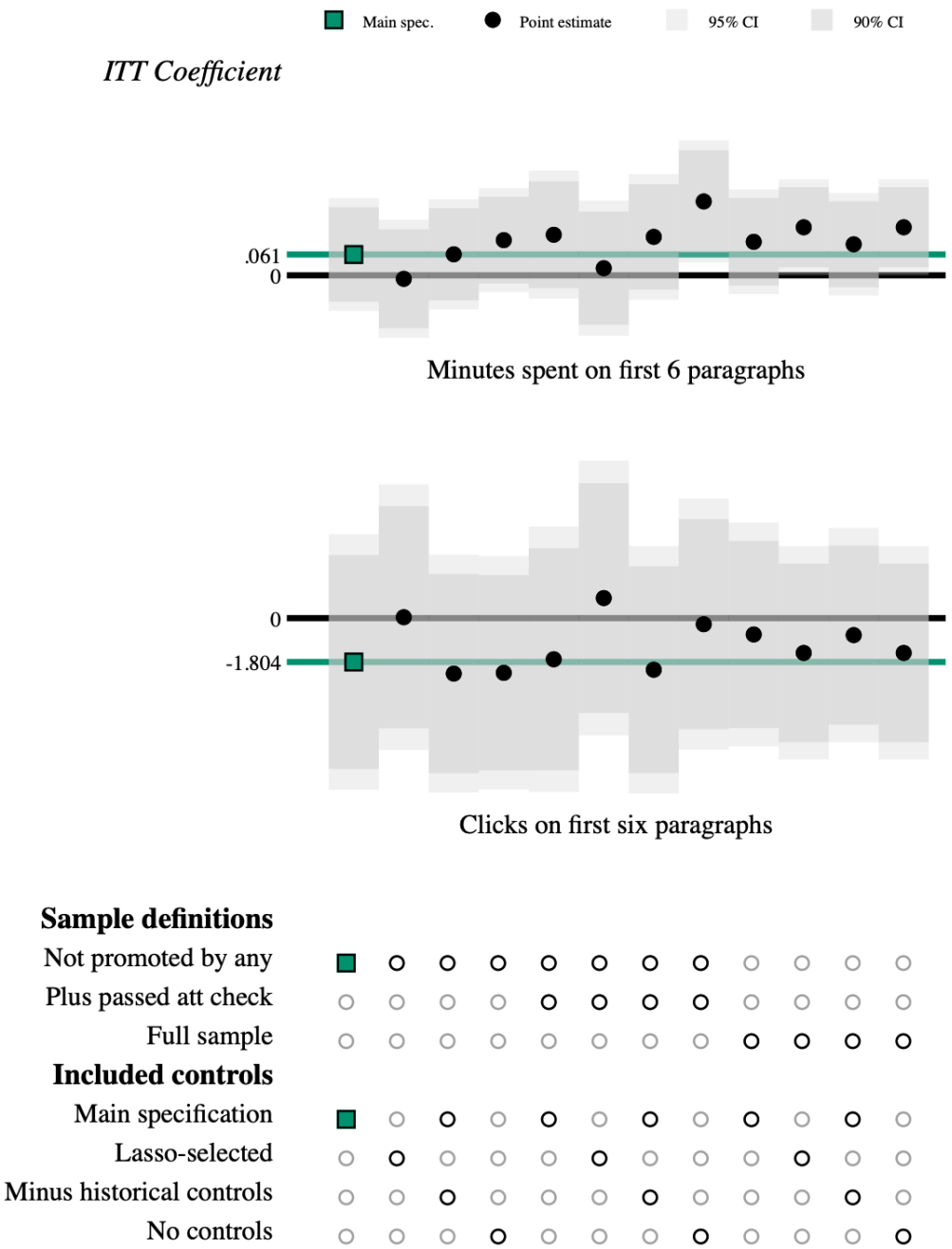
Note: This figure plots the treatment effect of seeing that the demographic-blind algorithm previously promoted mostly white men on perceived discrimination in the promotion experiment, relative to the other workers in the demographic-blind algorithm group who do not see the avatars, race, and gender 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 A16.
[Click to return to Section 4 (Robustness)]

Figure A25: Alternative specifications: Effects of seeing avatars in the algorithm arm on retention



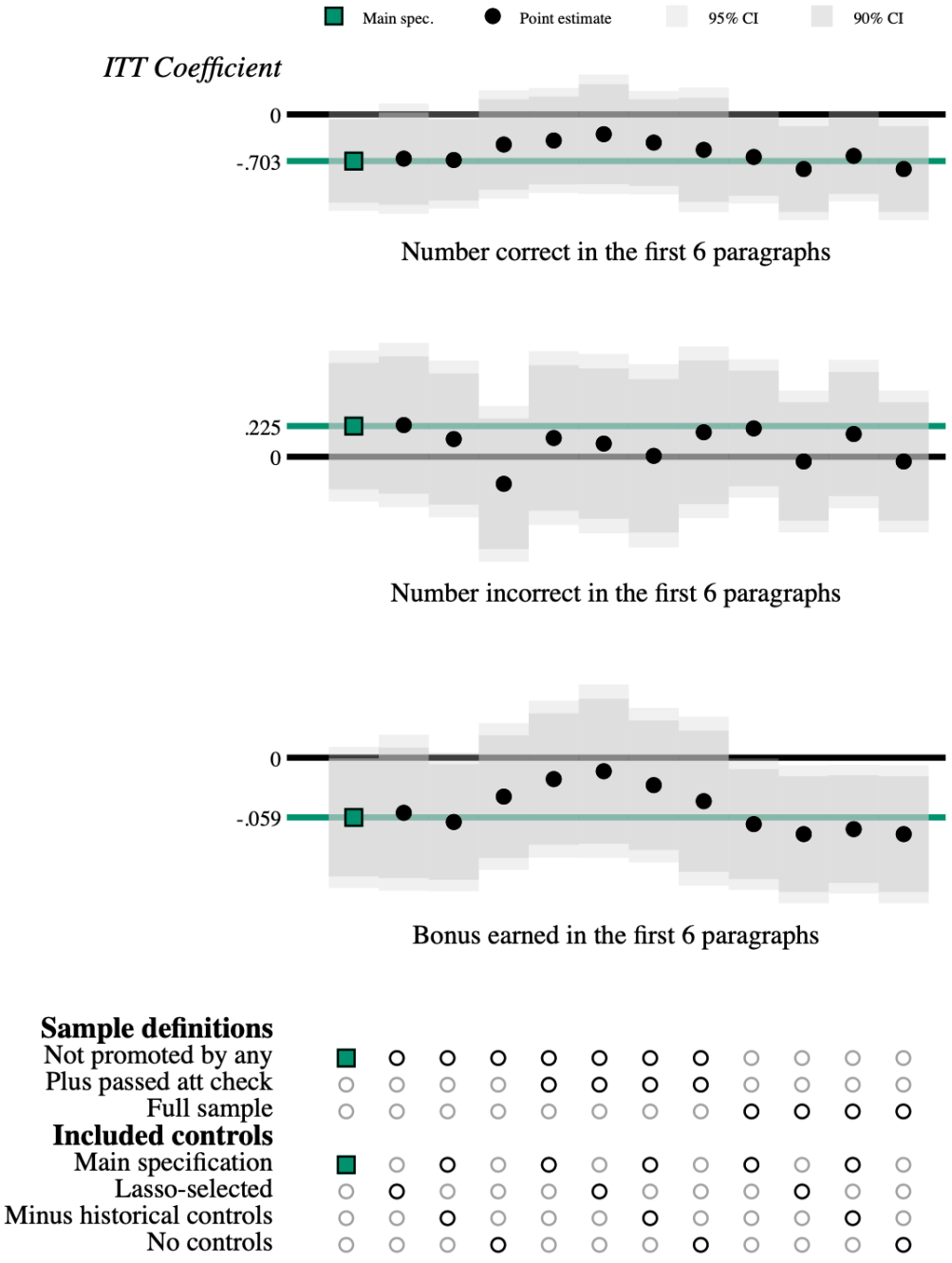
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on retention in the avatar arm of the promotion experiment. The outcomes and main specification are those in Table 3. Otherwise, the figure is analogous to Appendix Figure A24.
[Click to return to Section 4 (Robustness)]

Figure A26: Alternative specifications: Effects of seeing avatars in the algorithm arm on effort



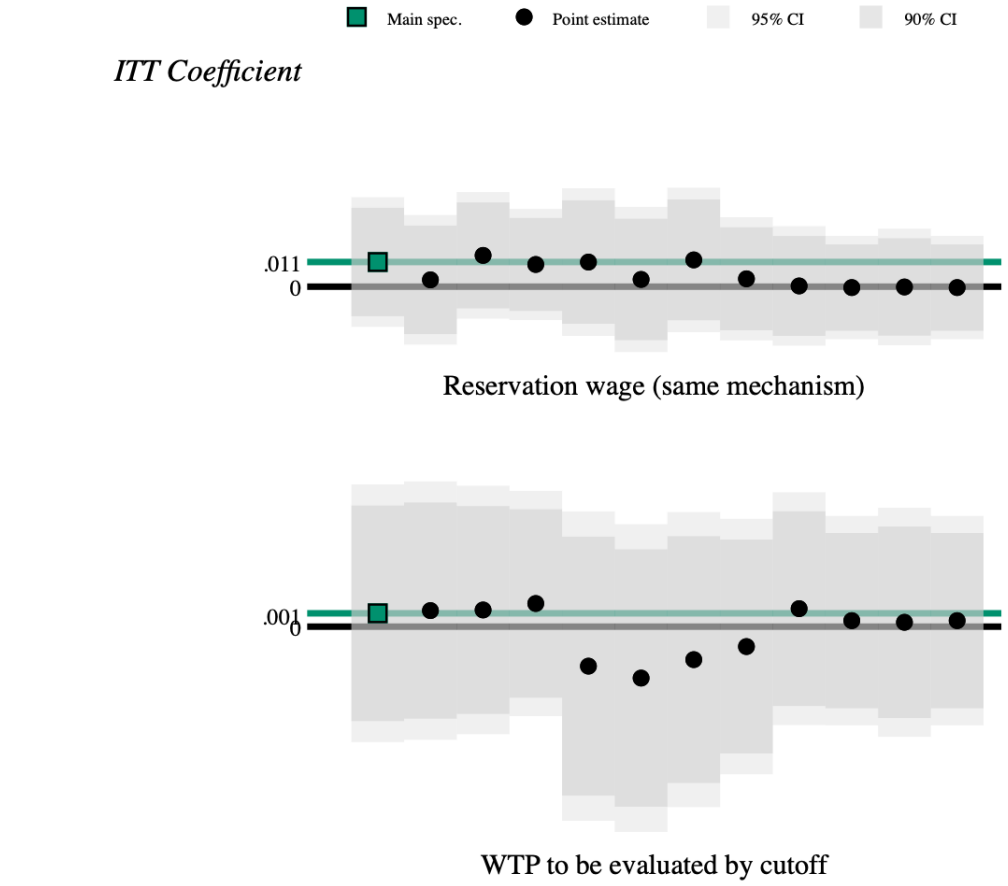
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on effort in the avatar arm of the promotion experiment. The outcomes and main specification are those in columns 1 and 2 of Table 4. Otherwise, the figure is analogous to Appendix Figure A24.
[Click to return to Section 4 (Robustness)]

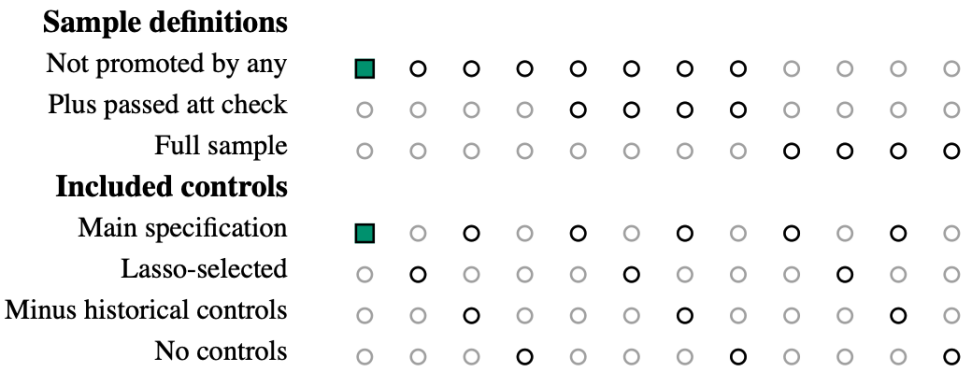
Figure A27: Alternative specifications: Effects of seeing avatars in the algorithm arm on performance



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 4. Otherwise, the figure is analogous to Appendix Figure A24.
[Click to return to Section 4 (Robustness)]

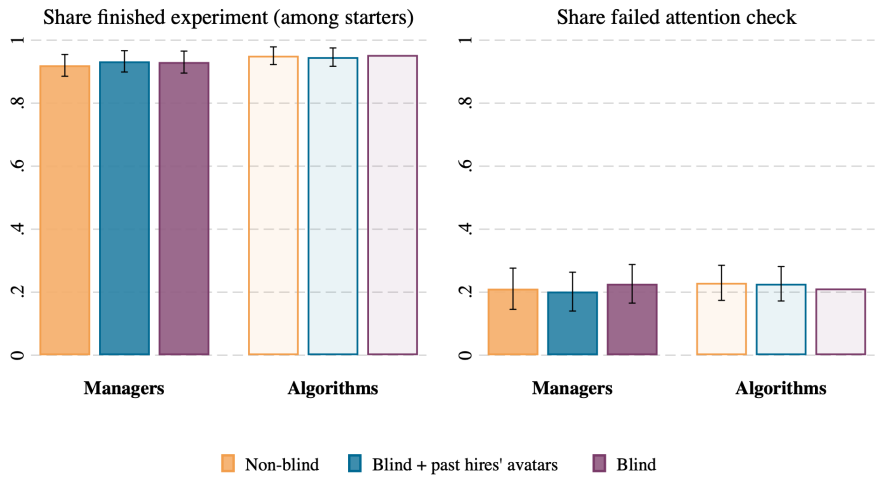
Figure A28: Alternative specifications: Effects of seeing avatars in the algorithm arm on future labor supply





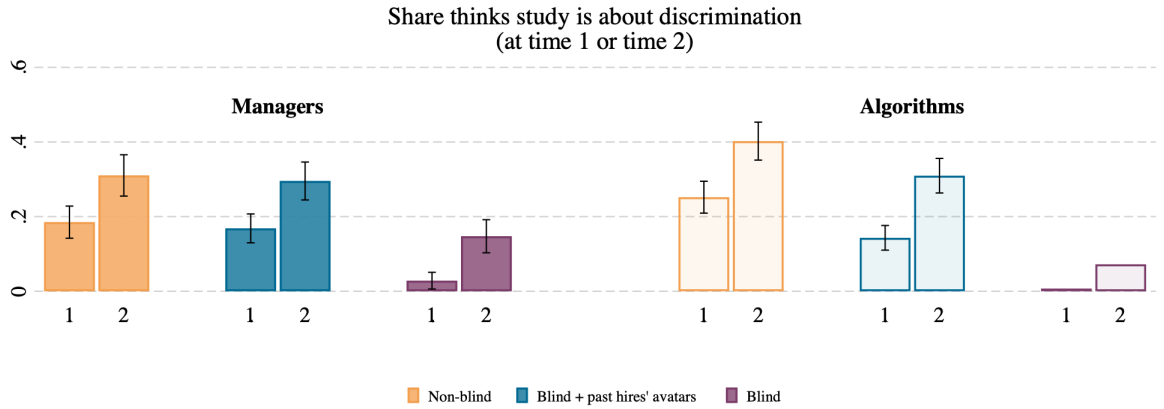
Note: This figure plots the treatment effect of seeing previously-promoted workers' avatars on reservation wages in the avatar arm of the promotion experiment. The outcomes and main specification are those in Table 5. Otherwise, the figure is analogous to Appendix Figure A24.
[Click to return to Section 4 (Robustness)]

Figure A29: Attrition and attention in the hiring experiment



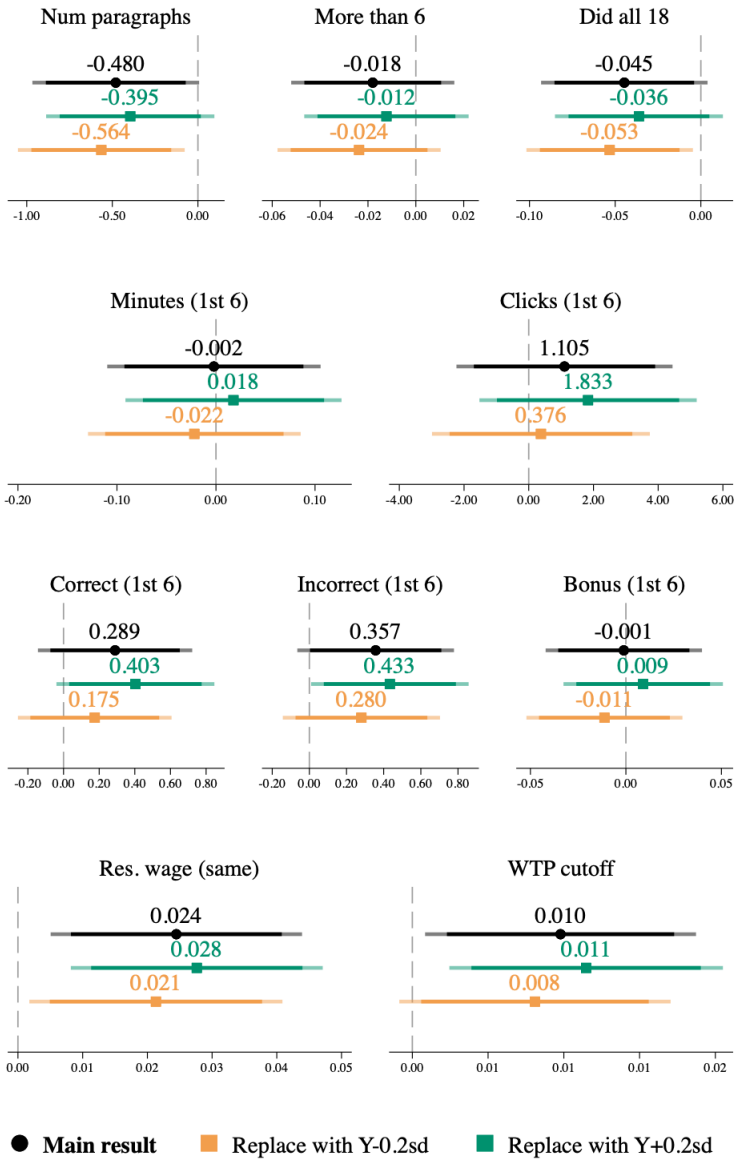
Note: This figure shows the fraction of workers in each arm of the hiring experiment who finished the experiment, among those who started it (on the left) and the share of the main sample that failed the attention check (on the right). The sample is restricted to workers who would not have been promoted 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.
[Click to return to Section 4 (Robustness)]

Figure A30: Knowledge of study topic in the hiring experiment



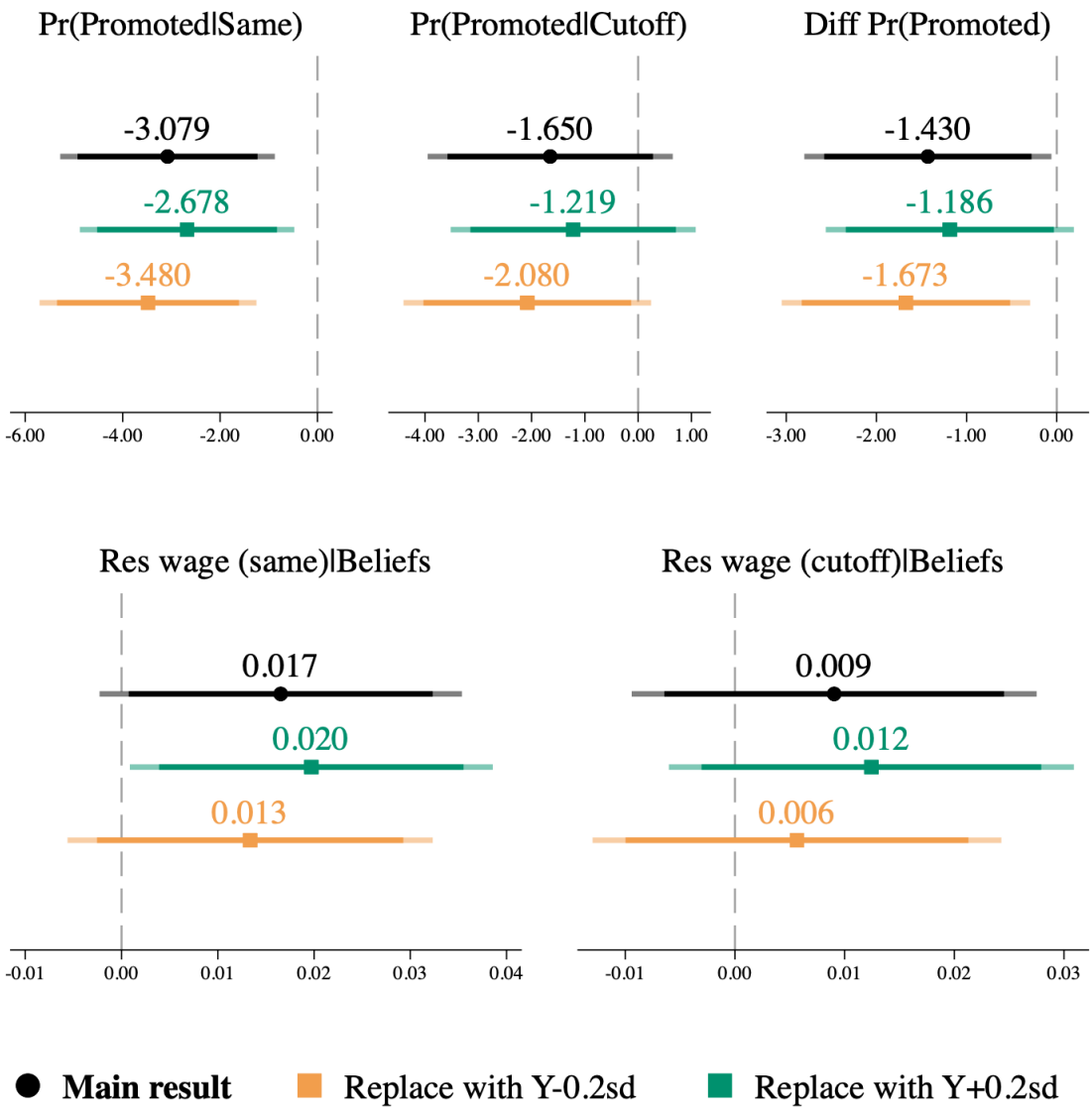
Note: This figure shows the fraction of workers in each arm of the hiring experiment who thought that the study was about discrimination the first time they were asked (after learning that they weren't hired and the reservation wage elicitations, before any survey questions) and the second (at the end 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.
[Click to return to Section 4 (Robustness)]

Figure A31: Accounting for possible demand effects in the manager arms of the promotion experiment: effort and future labor supply



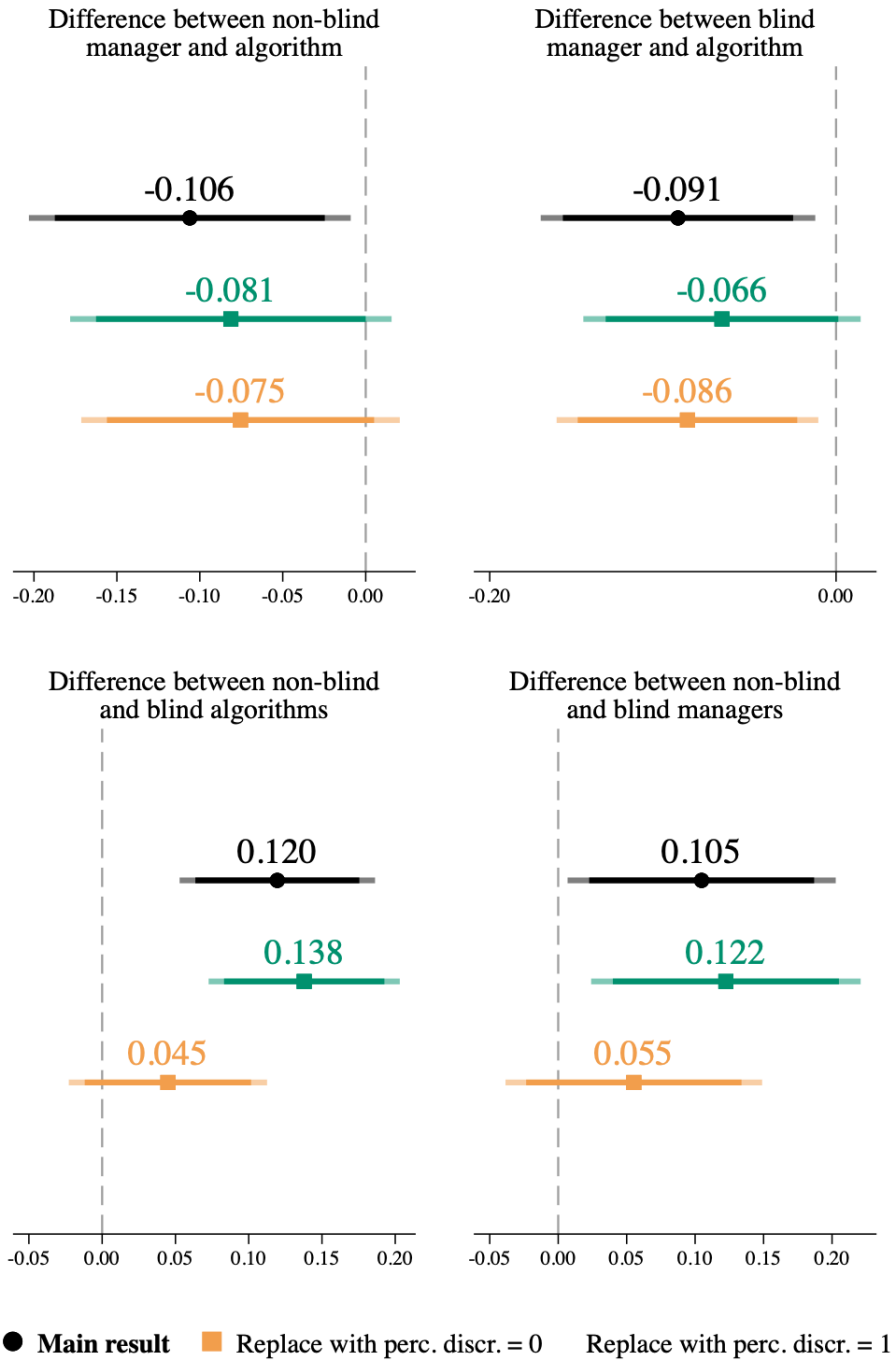
Note: This figure plots the treatment effects of being in the non-blind manager arm in the promotion experiment, relative to the demographic-blind manager arm, for the outcomes in Tables 3, 4, and 5, with adjustments to account for possible experimenter demand effects. These bounds replace the outcome 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 at the end of the study. 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.
[Click to return to Section 4 (Robustness)]

Figure A32: 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 Table 7 and Figure 10, with adjustments to account for possible experimenter demand effects. The figure is otherwise analogous to Appendix Figure A31.
[Click to return to Section 4 (Robustness)]

Figure A33: Accounting for possible demand effects in the hiring experiment

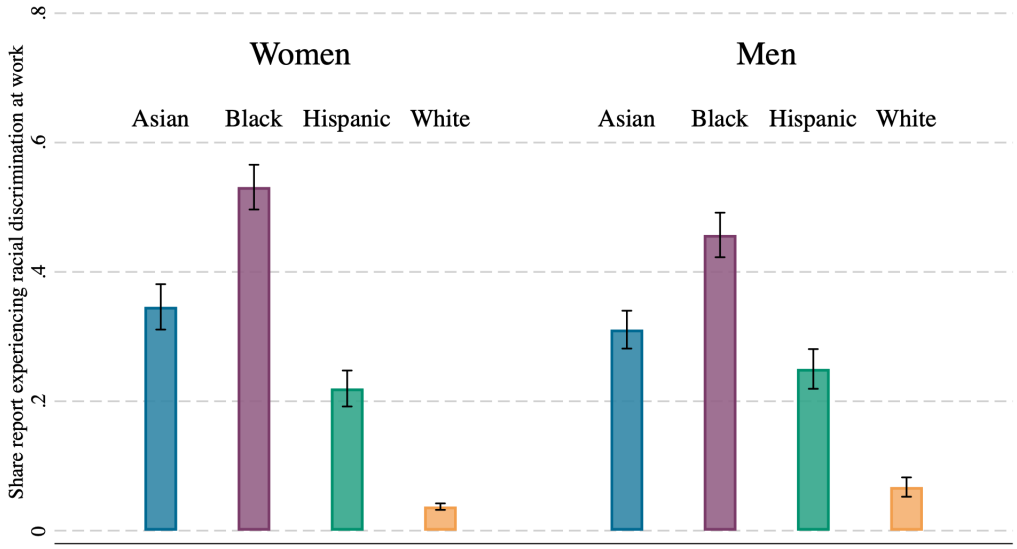


Note: This figure plots the differences in perceived discrimination between treatment arms in Figure 7, with adjustments to account for possible experimenter demand effects. These bounds replace the outcome variable (perceived discrimination) with a zero or one if workers thought that the study was about 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 7. 90 and 95 percent confidence intervals are calculated with standard errors robust to heteroskedasticity.

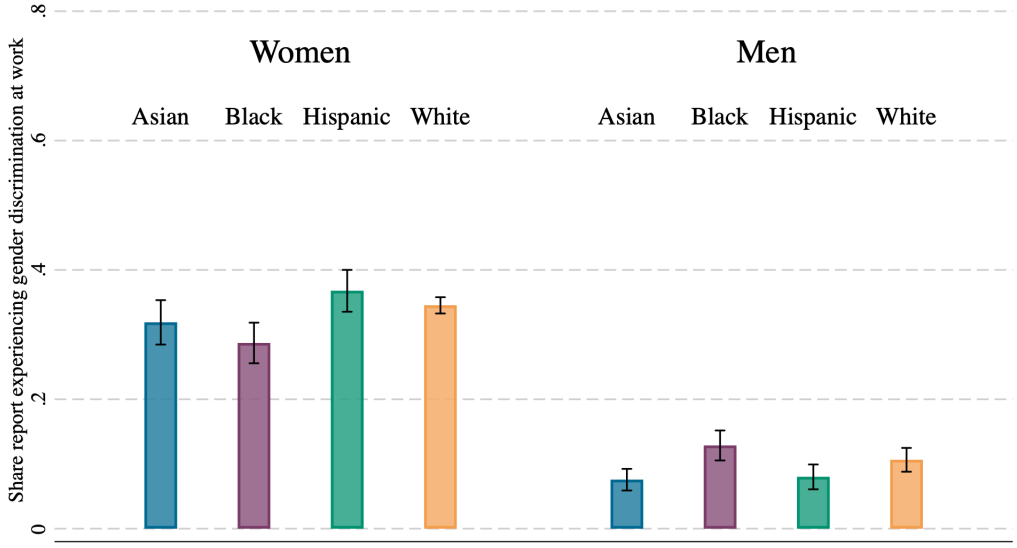
[Click to return to Section 4 (Robustness)]

Figure A34: Reported past experiences of discrimination at work in the experimental sample

Panel A: Perceived racial discrimination



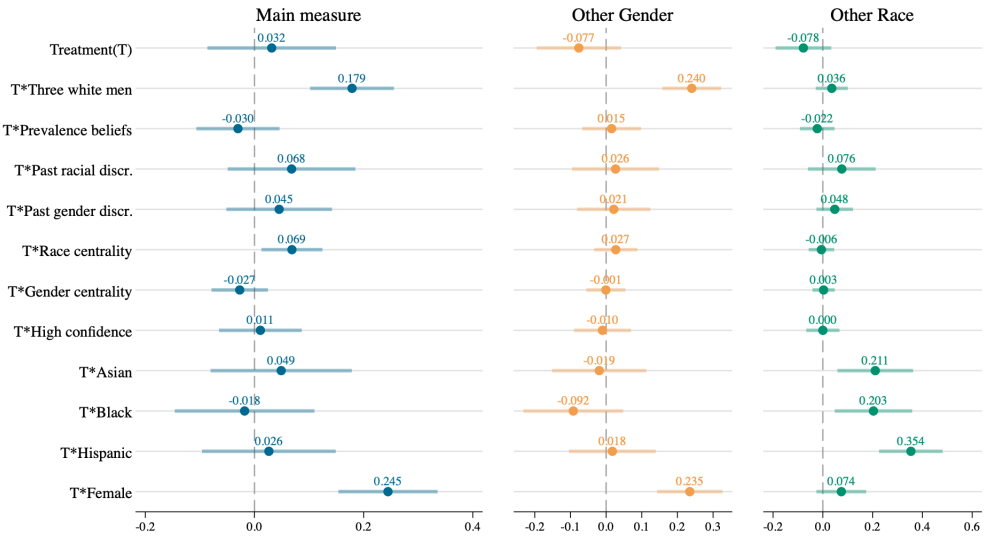
Panel B: Perceived gender discrimination



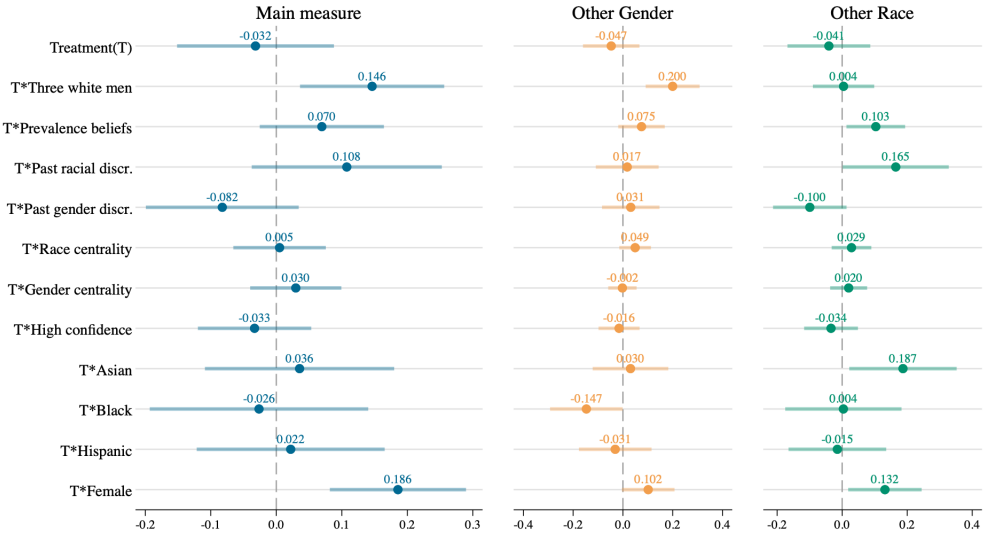
Note: The sample is 3,240 workers recruited for the promotion experiment on Prolific. The share of each group who reported having experienced discrimination at work in the past is plotted with standard errors in this sample indicated by black bars. Perceived discrimination includes during job search, 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 A35: All predictors of perceived discrimination in the promotion experiment

Panel A: Manager sample

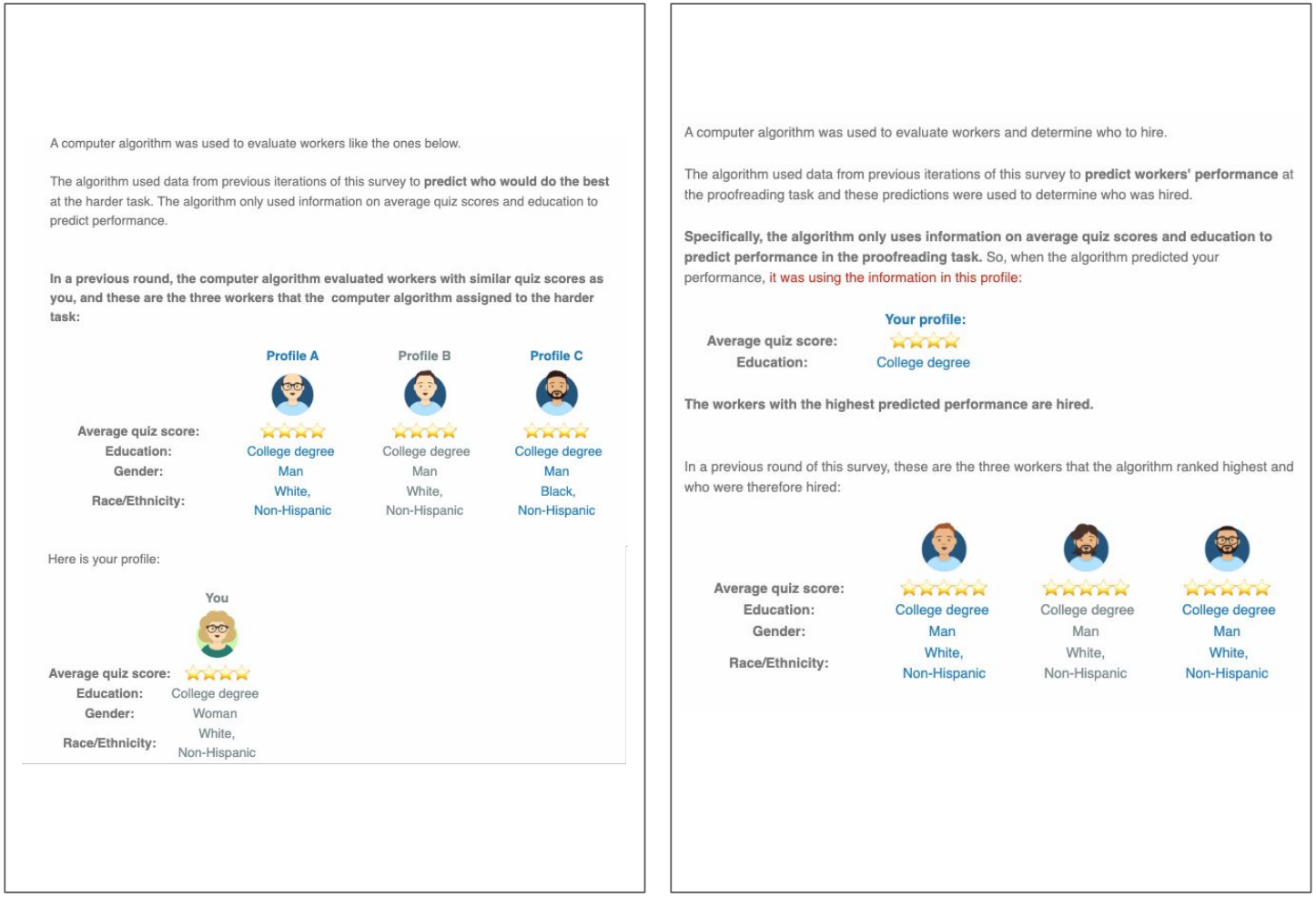


Panel B: Algorithm sample



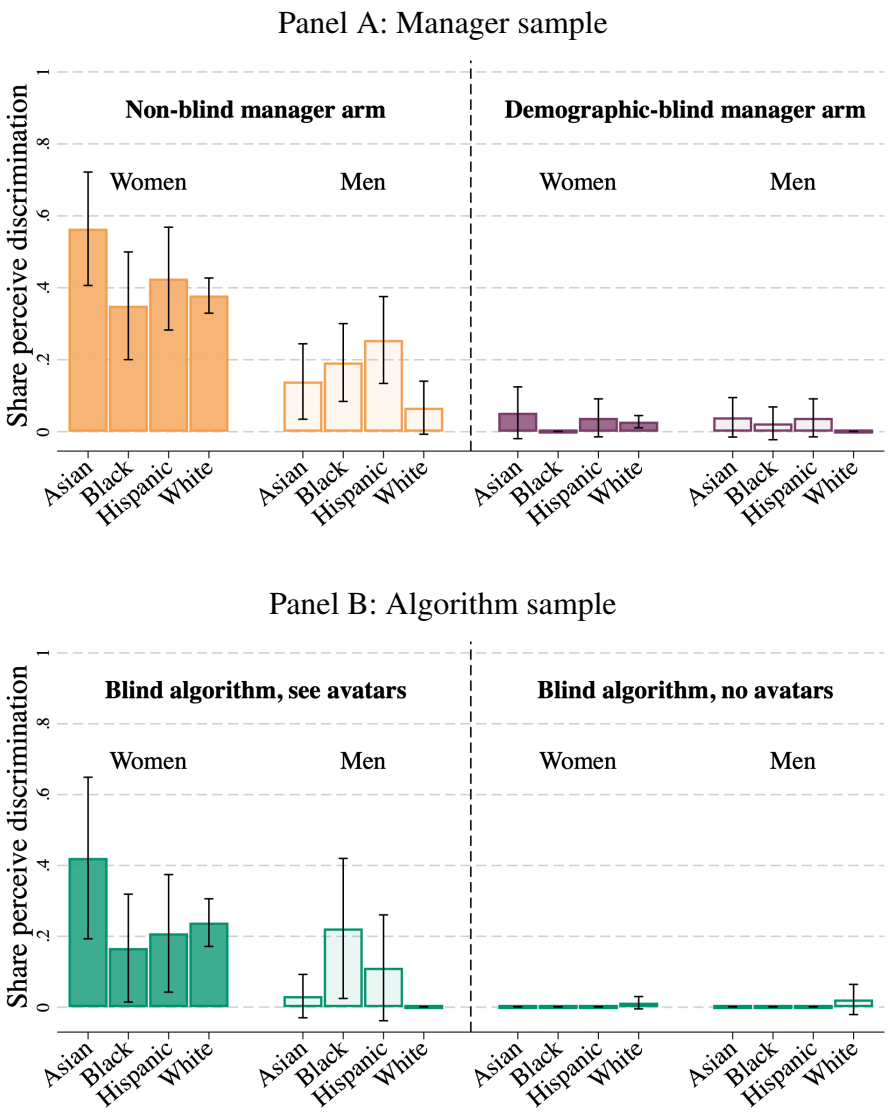
Note: This figure plots coefficients from a regression of the primary and secondary measures of perceived discrimination on an indicator for treatment, that indicator interacted with each of the worker characteristics listed on the left, the worker characteristics, and the control variables in the main 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 3.1 (Perceived Discrimination)]

Figure A36: Making information about decision-making inputs visually salient in the hiring experiment



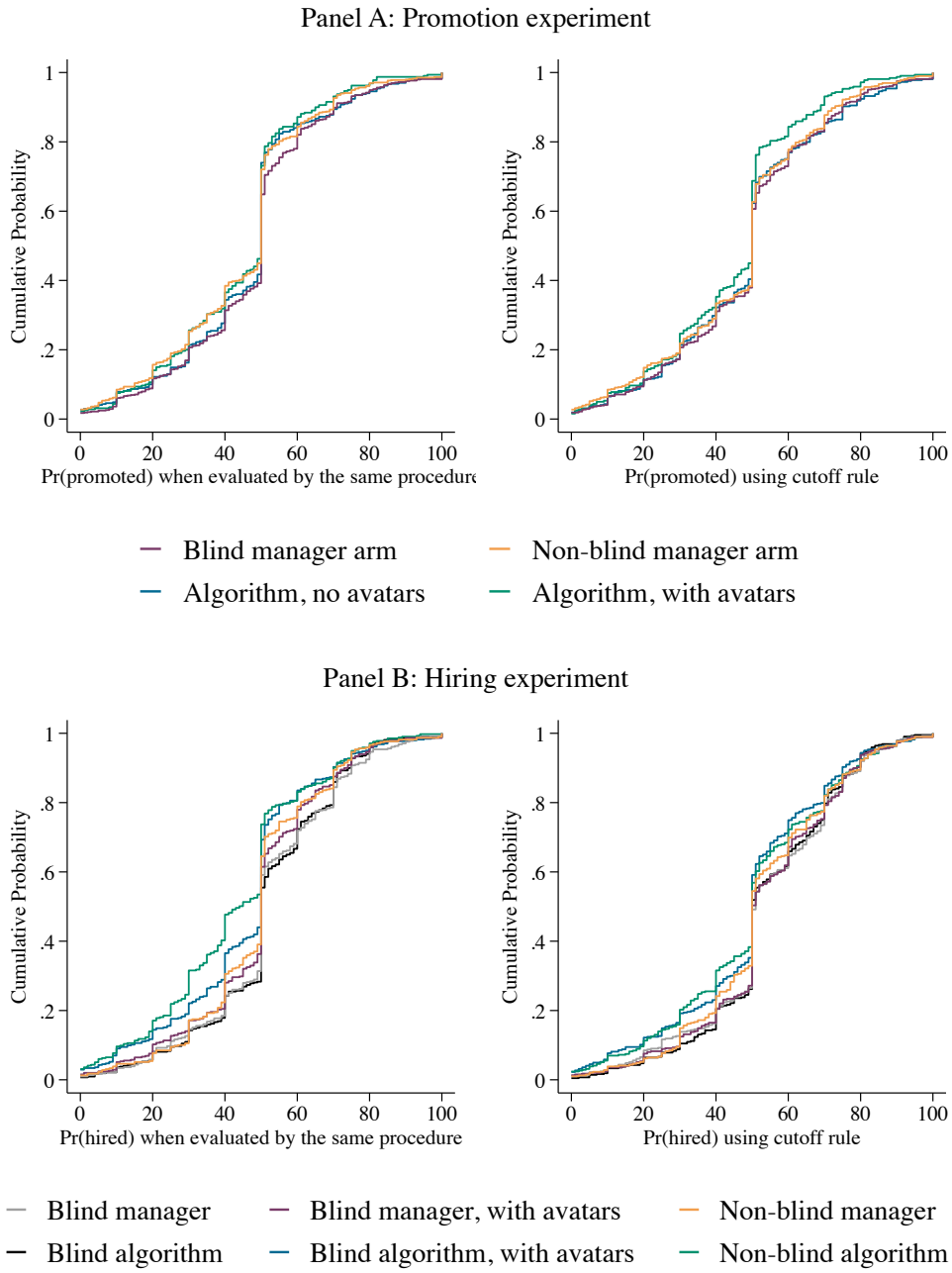
Note: This figure shows the only difference in implementation of the replicated treatment arms in the hiring experiment relative to the promotion experiment. In the promotion experiment, workers evaluated by a non-blind algorithm but who saw previously-promoted workers' avatars, race, and gender may 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 3.1 (Perceived Discrimination)]

Figure A37: 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 3.1 (Perceived Discrimination)]

Figure A38: CDFs of workers' beliefs about the likelihood of promotion and hiring in the future



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 (Future Labor Supply)]

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 of the promotion experiment

	All Workers	Women	Minority Men	White Men
Panel A: Free response, primary reason is discrimination				
Non-blind manager,	0.182***	0.199***	0.151***	0.059
see avatars	(0.016)	(0.020)	(0.031)	(0.041)
Control mean	0.010	0.014	0.000	0.000
Panel B: Yes, would be promoted if different race or gender				
Non-blind manager,	0.250***	0.294***	0.152***	0.053
see avatars	(0.023)	(0.028)	(0.054)	(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***	0.118***	0.052**	0.054
see avatars	(0.013)	(0.016)	(0.021)	(0.047)
Control mean	0.003	0.004	0.000	0.000
*Panel D: Members of my race * gender group underrepresented among promoted*				
Non-blind manager,	0.151***	0.206***	-0.028	0.015
see avatars	(0.026)	(0.032)	(0.062)	(0.044)
Control mean	0.386	0.378	0.538	0.020
Panel E: White men are overrepresented among promoted				
Non-blind manager,	0.155***	0.196***	0.023	0.329**
see avatars	(0.026)	(0.030)	(0.060)	(0.134)
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 Panel A of 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 3.1 (Perceived Discrimination)]

Table A6: 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 Panel B of 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 3.1 (Perceived Discrimination)]

Table A7: Effects on retention in the manager sample of the promotion experiment, 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	-0.654**	-0.041*	-0.046
	(0.315)	(0.021)	(0.032)
Non-blind manager arm*Minority women	-1.013*	-0.032	-0.120**
	(0.556)	(0.038)	(0.057)
Non-blind manager arm*Minority men	0.790	0.060	0.065
	(0.552)	(0.042)	(0.053)
p-values:			
White=minority, women	*0.562*	*0.823*	*0.246*
Men=women, minority	*0.020*	*0.098*	*0.016*
N	1293	1293	1293
Panel B: Gender heterogeneity			
Non-blind manager arm*Women	-0.751***	-0.038**	-0.066**
	(0.283)	(0.019)	(0.028)
Non-blind manager arm*Minority men	0.789	0.060	0.065
	(0.552)	(0.042)	(0.053)
p-values:			
Men=women	*0.011*	*0.029*	*0.026*
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 3, 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 3.2 (Retention and Performance)]

Table A8: Effects on effort and performance in the manager sample of the promotion experiment, gender and racial heterogeneity

	Minutes (1st 6)	Clicks (1st 6)	Correct (1st 6)	Incorrect (1st 6)	Bonus (1st 6)
	(1)	(2)	(3)	(4)	(5)
Panel A: Race and gender heterogeneity					
Non-blind manager arm*White women	-0.048	1.529	-0.074	0.176	-0.024
	(0.072)	(2.359)	(0.291)	(0.252)	(0.028)
Non-blind manager arm*Minority women	0.005	4.978	0.777	0.944**	0.033
	(0.116)	(4.354)	(0.519)	(0.394)	(0.048)
Non-blind manager arm*Minority men	0.125	-4.064	1.281***	0.110	0.066
	(0.121)	(2.535)	(0.468)	(0.522)	(0.044)
p-values:					
White=minority, women	*0.695*	*0.493*	*0.143*	*0.089*	*0.289*
Men=women, minority	*0.465*	*0.071*	*0.466*	*0.200*	*0.616*
N	1295	1295	1295	1295	1295
Panel B: Gender heterogeneity					
Non-blind manager arm*Women	-0.033	2.461	0.156	0.384*	-0.009
	(0.062)	(2.049)	(0.262)	(0.220)	(0.025)
Non-blind manager arm*Minority men	0.126	-4.046	1.286***	0.113	0.066
	(0.121)	(2.533)	(0.468)	(0.521)	(0.044)
p-values:					
Men=women	*0.226*	*0.045*	*0.035*	*0.628*	*0.135*
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 4, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A7. 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 3.2 (Retention and Performance)]

Table A9: Effects on future labor supply in the manager sample of the promotion experiment, gender and racial heterogeneity

	RW (Same)	WTP cutoff	Coop. task RW	WTP new manager
	(1)	(2)	(3)	(4)
Panel A: Race and gender heterogeneity				
Non-blind manager arm*White women	0.033**	0.011**	0.026	0.052**
	(0.013)	(0.006)	(0.019)	(0.026)
Non-blind manager arm*Minority women	-0.006	0.010	-0.019	0.116***
	(0.021)	(0.011)	(0.029)	(0.044)
Non-blind manager arm*Minority men	0.026	0.012	-0.011	-0.014
	(0.022)	(0.012)	(0.032)	(0.046)
p-values:				
White=minority, women	*0.110*	*0.894*	*0.188*	*0.201*
Men=women, minority	*0.292*	*0.885*	*0.838*	*0.043*
N	1247	1235	1256	957
Panel B: Gender heterogeneity				
Non-blind manager arm*Women	0.023**	0.011**	0.014	0.068***
	(0.011)	(0.005)	(0.016)	(0.023)
Non-blind manager arm*Minority men	0.026	0.012	-0.011	-0.013
	(0.022)	(0.012)	(0.032)	(0.046)
p-values:				
Men=women	*0.891*	*0.927*	*0.480*	*0.110*
N	1247	1235	1256	957
Control Mean, white women	0.253	-0.009	0.325	0.205
Control Mean, minority women	0.280	-0.019	0.349	0.218
Control Mean, minority men	0.280	-0.026	0.366	0.264

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 5, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A7. 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 (Future Labor Supply)]

Table A10: 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)
Panel A: Manager sample, promotion experiment					
Non-blind manager,	-0.075***	-0.046	-0.039	-0.060**	-0.035
see avatars	(0.027)	(0.028)	(0.028)	(0.028)	(0.026)
N	1380	1380	1380	1380	1389
Control mean	0.657	0.561	0.584	0.532	0.689
Panel B: Algorithm sample, promotion experiment					
See avatars	-0.016	0.024	0.026	-0.043	0.032
	(0.040)	(0.044)	(0.044)	(0.045)	(0.042)
N	688	688	688	688	691
Control mean	0.686	0.575	0.558	0.531	0.679
p-value, manager = algorithm	*0.215*	*0.171*	*0.204*	*0.739*	*0.164*

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 3. 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 (Future Labor Supply)]

Table A11: Effects on psychological well-being in the hiring experiment

	Percieved Discrimination	Good Affect	Self-efficacy
	(1)	(2)	(3)
Panel A: Manager sample, hiring experiment			
Non-blind manager	0.361***	-0.055	-0.022
or blind mgr + avatars	(0.020)	(0.042)	(0.046)
N	1246	1246	1246
Control mean	0.007	-0.010	-0.013
Panel B: Algorithm sample, hiring experiment			
Non-blind algorithm	0.529***	-0.062	0.031
or blind alg + avatars	(0.018)	(0.039)	(0.042)
N	1278	1278	1278
Control mean	0.002	-0.048	-0.097

This table reports treatment effects on perceived discrimination and psychological well-being in the hiring experiment. The outcome in the first column is the main measure of perceived discrimination. The outcome in column 2 is an index of psychological well-being and in column 3, self-efficacy in skills related to the proofreading task, both measured in standard deviations. The specifications are otherwise analogous to Table 8. 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 (Future Labor Supply)]

Table A12: 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	0.000	-0.004
	(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

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. Other than these changes in the sample, the specifications in each panel 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 4 (Robustness)]

Table A13: 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	0.028***	0.090***	0.103***
	(0.011)	(0.022)	(0.028)
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

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). The sample and specifications in each panel are otherwise 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 4 (Robustness)]

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: 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	0.25	0.29	*0.208*	0.27	0.28	*0.687*	0.31	0.29	*0.711*	0.32	0.36	*0.530*
Race:												
Asian	0.14	0.10	*0.126*	0.13	0.12	*0.763*	0.18	0.13	*0.269*	0.12	0.15	*0.500*
Black	0.13	0.14	*0.806*	0.13	0.13	*0.934*	0.10	0.17	*0.088*	0.15	0.18	*0.446*
White, Hispanic	0.14	0.14	*0.854*	0.16	0.15	*0.787*	0.15	0.10	*0.186*	0.12	0.15	*0.387*
White, Non-Hispanic	0.59	0.62	*0.479*	0.58	0.59	*0.732*	0.57	0.60	*0.578*	0.62	0.52	*0.099*
Married	0.51	0.52	*0.731*	0.53	0.54	*0.827*	0.52	0.57	*0.332*	0.65	0.52	*0.019*
Kids	0.47	0.43	*0.239*	0.41	0.44	*0.436*	0.38	0.49	*0.060*	0.44	0.41	*0.632*
Education:												
Less than high school	0.01	0.01	*0.796*	0.00	0.01	*0.244*	0.02	-0.00	*0.099*	0.01	0.02	*0.577*
High school graduate	0.15	0.14	*0.700*	0.12	0.14	*0.480*	0.12	0.15	*0.487*	0.09	0.11	*0.562*
Some college but no degree	0.25	0.28	*0.530*	0.33	0.27	*0.075*	0.24	0.32	*0.104*	0.25	0.25	*0.997*
2 year college degree	0.08	0.11	*0.346*	0.11	0.10	*0.605*	0.07	0.07	*0.905*	0.09	0.10	*0.894*
4 year college degree	0.36	0.33	*0.511*	0.31	0.39	*0.045*	0.44	0.36	*0.179*	0.38	0.39	*0.868*
Professional or Masters degree	0.13	0.12	*0.867*	0.11	0.08	*0.242*	0.10	0.09	*0.849*	0.15	0.12	*0.402*
Doctorate	0.02	0.01	*0.556*	0.01	0.01	*0.684*	0.01	0.01	*0.805*	0.02	0.01	*0.426*
Income:												
Less than $20,000	0.13	0.15	*0.416*	0.14	0.13	*0.698*	0.14	0.15	*0.779*	0.09	0.17	*0.046*
$20,000-$40,000	0.21	0.21	*0.948*	0.22	0.25	*0.340*	0.26	0.19	*0.116*	0.18	0.22	*0.413*
$40,000-$60,000	0.19	0.20	*0.814*	0.17	0.17	*0.986*	0.18	0.22	*0.303*	0.21	0.13	*0.079*
$60,000-$80,000	0.15	0.15	*0.825*	0.14	0.15	*0.785*	0.16	0.16	*0.961*	0.16	0.13	*0.345*
$80,000-$100,000	0.09	0.12	*0.275*	0.14	0.11	*0.225*	0.08	0.09	*0.649*	0.12	0.09	*0.386*
$100,000-$120,000	0.10	0.05	*0.011*	0.06	0.08	*0.456*	0.05	0.05	*0.941*	0.04	0.09	*0.131*
$120,000-$140,000	0.03	0.01	*0.062*	0.05	0.04	*0.451*	0.04	0.03	*0.617*	0.04	0.06	*0.434*
$140,000-$160,000	0.03	0.05	*0.241*	0.03	0.03	*0.765*	0.03	0.04	*0.527*	0.06	0.04	*0.284*
More than $160,000	0.06	0.05	*0.590*	0.05	0.05	*0.722*	0.07	0.07	*0.953*	0.09	0.07	*0.529*
Age:												
18-24	0.14	0.18	*0.134*	0.17	0.17	*0.814*	0.20	0.15	*0.196*	0.19	0.17	*0.592*
25-34	0.31	0.33	*0.570*	0.31	0.31	*0.933*	0.30	0.36	*0.258*	0.34	0.31	*0.505*
35-44	0.24	0.20	*0.145*	0.22	0.23	*0.858*	0.26	0.19	*0.116*	0.17	0.22	*0.311*
45-54	0.14	0.12	*0.515*	0.17	0.14	*0.268*	0.12	0.12	*0.911*	0.15	0.15	*0.914*
55-64	0.10	0.12	*0.358*	0.08	0.12	*0.157*	0.05	0.14	*0.003*	0.11	0.12	*0.763*
65 or older	0.07	0.05	*0.227*	0.05	0.04	*0.729*	0.06	0.05	*0.536*	0.04	0.04	*0.879*
Employment:												
Currently employed outside Prolific	0.64	0.65	*0.886*	0.66	0.66	*0.892*	0.71	0.73	*0.672*	0.74	0.64	*0.052*
Ever employed outside Prolific	0.93	0.97	*0.023*	0.96	0.96	*0.826*	0.97	0.98	*0.870*	0.99	0.94	*0.020*
Satisfied with current employer	0.80	0.78	*0.598*	0.75	0.79	*0.381*	0.74	0.77	*0.666*	0.80	0.75	*0.319*
Satisfied with most recent employer	0.56	0.66	*0.149*	0.58	0.69	*0.093*	0.61	0.72	*0.308*	0.66	0.64	*0.886*
N	377	318		475	219		129	191		254	117	

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 A16: 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, see avatars	-0.101***	-0.115***	-0.113**	0.092
	(0.024)	(0.028)	(0.055)	(0.099)
Control mean	0.414	0.410	0.428	0.408
Panel B: Free response, first answer is higher quiz score				
Non-blind manager, see avatars	-0.059**	-0.078**	-0.008	-0.013
	(0.026)	(0.031)	(0.063)	(0.105)
Control mean	0.484	0.488	0.476	0.469
Panel C: Free response, first answer is it was random				
Non-blind manager, see avatars	-0.003	0.003	-0.003	-0.086
	(0.010)	(0.013)	(0.020)	(0.061)
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 in Panel A of 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 3.1 (Perceived Discrimination)]

Table A17: 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 A16. Specifications are otherwise the same as in Panel B of 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 3.1 (Perceived Discrimination)]

Table A18: Effects on psychological outcomes in the promotion experiment

	Total good affect (SD)	Self-efficacy index (SD)	Very satisfied with work
	(1)	(2)	(3)
Panel A: Manager sample, promotion experiment			
Non-blind manager, see avatars	0.004	-0.005	-0.037
	(0.035)	(0.039)	(0.028)
N	1389	1382	1389
Control mean	0.011	-0.013	0.549
Panel B: Algorithm sample, promotion experiment			
See avatars	0.073	-0.105	-0.007
	(0.057)	(0.065)	(0.046)
N	691	689	691
Control mean	-0.039	-0.037	0.536
p-value, manager = algorithm	*0.289*	*0.176*	*0.568*

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 3.2 (Retention and Performance)]

Table A19: Effects on psychological outcomes in the manager arms of the promotion experiment, 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	0.014	-0.011
	(0.047)	(0.052)	(0.038)
Non-blind manager arm*Minority women	-0.064	-0.044	-0.051
	(0.079)	(0.093)	(0.061)
Non-blind manager arm*Minority men	0.105	-0.010	-0.056
	(0.070)	(0.083)	(0.058)
p-values:			
White=minority, women	*0.505*	*0.583*	*0.573*
Men=women, minority	*0.109*	*0.784*	*0.957*
N	1295	1288	1295
Panel B: Gender heterogeneity			
Non-blind manager arm*Women	-0.020	-0.002	-0.022
	(0.041)	(0.047)	(0.033)
Non-blind manager arm*Minority men	0.104	-0.010	-0.056
	(0.070)	(0.083)	(0.058)
p-values:			
Men=women	*0.124*	*0.933*	*0.609*
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 A18, but white men are additionally dropped from the sample. The specification is the same as Appendix Table A7. 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 3.2 (Retention and Performance)]

B Manager recruitment and manager task

Managers were also 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 2, 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 hiring experiment (360 in the promotion 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. Thus, managers saw workers in the same quiz-score quintile groups in both the historical and main samples (workers were told "your manager evaluated workers with similar quiz scores as you and these are the three they assigned to the harder task"). Managers also saw the same information about workers in the historical and main samples—those who had been randomly assigned to see avatars, race, and

gender in the historical sample also saw these characteristics in the main sample, and vice versa.

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 to promote 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, 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 brought 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. Again, 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, with the same priorities described above) from the group of managers that had not been assigned to evaluate workers initially. This continued until all workers had been evaluated. 20 managers participated in the hiring experiment.

In one quiz-score quintile group of 120 workers in the promotion experiment, I ran out of managers who could be paired with the group after all of the eligible managers did not return. 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 quiz-score quintile 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, regardless). 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. In other words, each of the categories of quiz-score quintiles and education that the manager saw in the demographic-blind case are used as predictors. 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 measure of performance that the algorithm predicted was 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 or discretionary effort. 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: $E[Y|\hat{Y},R,G] = E[Y|\hat{Y}]$. In other words, 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 (and thus, mechanically, fewer women).

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 overrepresented 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—one key to ensuring an unbiased algorithm. 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 from both the easier and harder proofreading jobs.

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 directly 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 for clarity.*

- 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"*

  "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 **$0.45** for each high-quality easy paragraph; **$0.90** 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."*

* *In the demographic-blind algorithm arms, regardless of whether they see previous hires' avatars, race, and gender: "the same algorithm as was used to make today's hiring decisions will use your average quiz score and education to predict your performance and determine whether you would be hired for the proofreading task. The outcome may be different from today (you might be hired this time) because the algorithm would be evaluating a different set of workers."*
* *In the non-blind algorithm arm: "the same algorithm as was used to make today's hiring decisions will use your average quiz score, education, race, and gender to predict your performance and determine whether you would be hired for the proofreading task. The outcome may be different from today (you might be hired this time) because the algorithm would be evaluating 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 **$0.90** 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 A2).

- **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 **$0.95** 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):

  - I am capable of doing the harder proofreading job well
  - I would have liked a chance to do the harder proofreading task
  - I am confident in my ability to work under pressure
  - I cam capable of doing the easier proofreading job well
  - I did a good job on the proofreading task today
  - 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):

  - Written communication
  - Oral communication
  - Problem solving
  - Numeracy
  - Motivation
  - 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 A18, 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):
  - I want to complete more surveys for this employer
  - I want to complete more surveys with tasks that are assigned in this way
  - I want to complete more proofreading tasks of this level of difficulty
  - I want to complete more difficult proofreading tasks.

  The outcomes in columns 1-4 of Appendix Table A10 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):
  - How satisfied are you on the whole with the work that you were offered in this survey?
  - How satisfied are you on the whole with the work that is available on Prolific?
  - 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.

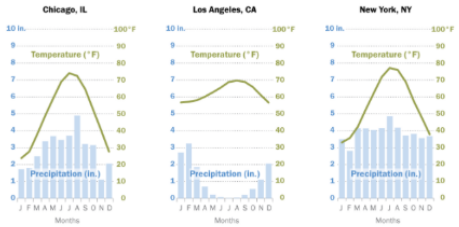
  - Which of these is not a word or phrase? [a lot, *alot*, allot]
  - You've probably heard the phrase "i before e except after c" ...but which of these words defies this rule? [species, science, policies, *all of them*]
  - A plural subject needs [a singular verb, *a plural verb*]
  - When two singular 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 charac-

teristic 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?

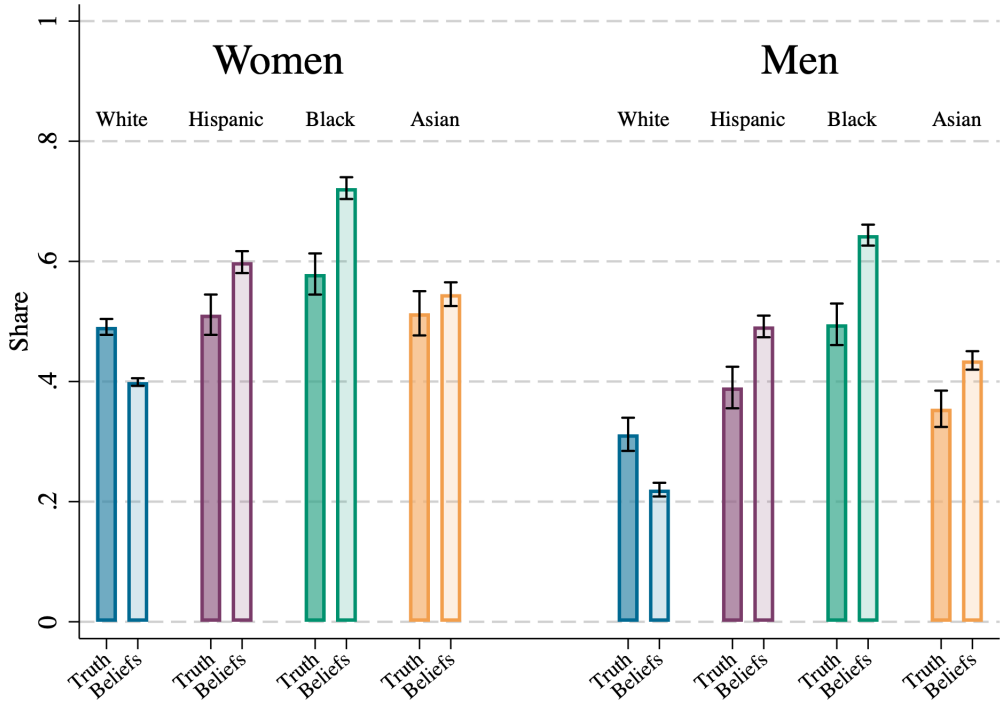
After each question, they could answer "Yes," "No," or "It's impossible to know." If they answered yes, they were asked if they thought they were discriminated against based on gender, race, age or something else, and to select all that apply, including an option to write in "other." Finally, they were 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 × 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 2.5, an instrumental-variables 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 is careful to eliminate 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.

As discussed above, 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, I focus on the ITT estimates in the paper. These are a lower bound on the average effects of perceived discrimination in the case that everyone perceives some discrimination. Here, for interested readers, I present the results from an IV specification that instruments the main measure of perceived discrimination with treatment assignment in Appendix Tables E1-E5. 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	-1.562*	-0.059	-0.145*
	(0.804)	(0.056)	(0.080)
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 3 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Table 2). 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 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	-0.007	3.593	0.941	1.160*	-0.004
	(0.176)	(5.469)	(0.708)	(0.693)	(0.067)
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 4 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Table 2). 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 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	0.078**	0.031**
	(0.031)	(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 5 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Table 2). 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.

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	-10.000***	-5.357	-4.643**
	(3.603)	(3.754)	(2.236)
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 7 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Table 2). Outcomes and control variables are otherwise the same as Table 7. 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	0.033	0.133**	0.159*	-0.976
	(0.045)	(0.059)	(0.090)	(0.735)
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 6 by dividing them by the first-stage ITT effects on the main measure of perceived discrimination (in column 1 of Table 2). 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.

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	-0.928*	-0.036	-0.107**
	(0.478)	(0.034)	(0.048)
Non-blind manager arm*Minority Men	1.419**	0.098**	0.112*
	(0.611)	(0.045)	(0.059)
Non-blind manager arm*Exp Past Discr	-0.539	0.005	-0.066
	(0.486)	(0.034)	(0.049)
Non-blind manager arm*High Discr Beliefs	0.259	-0.011	0.039
	(0.483)	(0.033)	(0.049)
Non-blind manager arm*High Confidence	0.600	0.001	0.108**
	(0.478)	(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 Panel A of Table 3, 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 (1st 6)	Clicks (1st 6)	Correct (1st 6)	Incorrect (1st 6)	Bonus (1st 6)
	(1)	(2)	(3)	(4)	(5)
Non-blind manager arm	-0.042	3.088	0.056	0.694**	-0.022
	(0.105)	(3.124)	(0.435)	(0.338)	(0.040)
Non-blind manager arm*Minority Men	0.167	-6.076*	1.124**	-0.191	0.074
	(0.132)	(3.301)	(0.541)	(0.553)	(0.051)
Non-blind manager arm*Exp Past Discr	0.104	1.375	0.443	0.133	0.015
	(0.109)	(3.607)	(0.447)	(0.396)	(0.043)
Non-blind manager arm*High Discr Beliefs	-0.089	-1.146	-0.394	-0.462	0.015
	(0.109)	(3.446)	(0.454)	(0.419)	(0.043)
Non-blind manager arm*High Confidence	0.005	-1.343	0.155	-0.287	-0.007
	(0.106)	(3.426)	(0.442)	(0.389)	(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 Panel A of 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 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	-0.553*	-0.018	-0.055*
	(0.295)	(0.021)	(0.030)
Non-blind manager arm*Not employed	-0.343	-0.018	-0.025
	(0.407)	(0.028)	(0.041)
p-values:			
Employed = not employed	*0.663*	*0.983*	*0.535*
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 Panel A of Table 3. 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 (1st 6)	Clicks (1st 6)	Correct (1st 6)	Incorrect (1st 6)	Bonus (1st 6)
	(1)	(2)	(3)	(4)	(5)
Non-blind manager arm*Employed	0.028	1.374	0.491*	0.310	0.004
	(0.067)	(2.114)	(0.270)	(0.263)	(0.026)
Non-blind manager arm*Not employed	-0.060	0.637	-0.082	0.462	-0.011
	(0.087)	(3.064)	(0.370)	(0.316)	(0.035)
p-values:					
Employed = not employed	*0.408*	*0.846*	*0.206*	*0.692*	*0.731*
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 Panel A of 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.

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	0.018	0.014	0.006	0.065
	(0.019)	(0.009)	(0.028)	(0.041)
Non-blind manager arm*Minority Men	0.007	0.002	-0.026	-0.074
	(0.025)	(0.013)	(0.035)	(0.051)
Non-blind manager arm*Exp Past Discr	0.017	0.004	0.004	0.063*
	(0.020)	(0.009)	(0.028)	(0.038)
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	-0.018	-0.007	0.008	0.008
	(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 Panel A of Table 5, 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	0.027**	0.011**	0.020	0.049**
	(0.012)	(0.006)	(0.017)	(0.025)
Non-blind manager arm*Not employed	0.019	0.007	-0.008	0.034
	(0.016)	(0.008)	(0.024)	(0.031)
p-values:				
Employed = not employed	*0.682*	*0.612*	*0.328*	*0.702*
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 Panel A of Table 5. 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 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 A15). 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.

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. I did not anticipate the treatment arms having such different rates of this, given the lack of differences across arms in the promotion experiment. 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 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 exper-

iment) 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.

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 4.2 however, this does not affect the main conclusions. Estimates without these controls are in Appendix Figures A14-A28.

De-emphasis of IV specification. Section 2.5 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.

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 dis-

crimination 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 A6 (Panel A) and Appendix Tables A16 and A17 (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 A6 (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 A13 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 A30.

I Implicit measures of perceived discrimination

The measures of perceived discrimination discussed in the paper are all explicit: they require the worker to think about or describe an experience of discrimination (disparate treatment or impact based on race or gender, even if not the word discrimination directly as in the main measure of perceived discrimination). I also observe implicit measures of discrimination: workers' beliefs about whether they in particular (or workers from different demographic groups in general) would be held to a higher or lower standard (in terms of quiz scores) to be promoted than an otherwise equivalent worker of a different race or gender. This is measured by asking workers about the number of stars (representing quiz-score quintiles) they think they and other workers would have needed on their profiles to be promoted. These are only asked in the promotion experiment, in order to shorten the experimental survey in the hiring experiment, given the noisy, small treatment effects in the results described below.

They were asked how many stars *they themselves* think they needed to be promoted right after the open-ended question that underlies the main measure of perceived discrimination. Treatment differences in the number of stars workers think they needed by race and gender are indicative of perceived discrimination (if women or racial minorities in the non-blind manager group think they need more stars than those in the demographic-blind manager group, for example), but this cannot be used to identify individuals who thought that they were discriminated against as one can with the explicit measures used in the text. It does, however, allow for observation of whether white and male workers are aware of any "positive discrimination" they experience (which they are unlikely to bring up in the more explicit measures, which identify only "negative discrimination").

Appendix Table I1, Panel A (below) shows differences in the number of *additional* stars workers think they would have needed to be promoted by treatment status in the manager sample. The specification is the same as in the main analysis. Consistent with perceived racial discrimination, racial minority women and men think they need more stars in the non-blind manager group than in the demographic-blind manager group, and white men think they need fewer, but none of these treatment effects are particularly large in magnitude compared to the control mean and only the effect for racial minority men is even marginally statistically significantly different from zero. Appendix Table I2, Panel A shows the analogous estimate in the algorithm sample. Here, racial minority women on average think they need almost an additional half of a star to be promoted in the arm where they can see previously-promoted workers' avatars, race, and gender relative to when they cannot. White men think they need 0.4 *fewer* additional stars when they see that primarily white men were previously promoted than when they don't, although this is not statistically significant.

At the very end of the survey, a "vignette-style" series of elicitations asked workers to also con-

sider that they were four different imaginary workers, and asked them how many stars they thought each of those workers would have needed on their profiles in order to be promoted. Comparing the differences between the number of stars workers think they and others needed provides additional information about perceived discrimination and can also be used to generate indicators for whether individual workers perceive some type of discrimination – needing more stars or fewer stars than an otherwise similar worker with a different race or gender.

Workers see more information in these imaginary-worker profiles that in those used throughout the rest of the survey: they see avatars, race, gender, education, income buckets, and family status (married, kids), in order to make it more realistic that they might be these workers. The four imaginary people that workers are asked to consider are constructed so that they were similar to the worker themselves other than in race and gender, but there was other variation to disguise the purpose of the exercise as much as possible. Income and education were randomly chosen among the options within one category of the workers' own. The four imaginary people had a 60 percent chance of having the same [*married*, *kids*] pair as the worker themselves, about an 18 percent chance of having the same *married* value but the opposite *kids* value and about an 18 percent change of the opposite, and about a 4 percent chance of having the opposite value for both. Avatars were randomly chosen from the set of real participants' chosen avatars who had the race and gender of the imaginary person. The race and gender of the four imaginary people were as follows:

1. **Different race, same gender (1)**
    - Men saw men and women see women; Non-binary respondents saw a profile with the binary gender the worker had reported to Prolific which allowed them to be screened into the sample
    - Workers who were white and non-Hispanic see profiles that were Black and non-Hispanic
    - Workers who were not white and non-Hispanic see profiles that were white and non-Hispanic
2. **Different race, same gender (2)**
    - Gender was the same as above
    - Workers who were white and non-Hispanic see profiles that were Asian
    - Workers who were white and Hispanic or Asian see profiles that were Black and non-Hispanic
    - Workers who were Black see profiles that where white and Hispanic

3. **Same race, different gender**

    - Men saw women and women see men; Non-binary respondents see a profile with the opposite of the binary gender the worker had reported to Prolific which allowed them to be screened into the sample
    - Workers who identified as multiracial or other race see profiles that had the race that they had reported to Prolific that allowed them to screen into the sample (white and non-Hispanic, white and Hispanic, Black, or Asian); everyone who identified as one of those four main groups saw a profile with the same race.

4. **Same race, same gender**

    - Race as in (3) and gender as in (1) and (2)

Appendix Tables I1 and I2, Panels B-E show differences in whether workers thought they needed more or fewer stars than otherwise similar workers, by treatment group among the manager sample. The specifications are the same as in the main analysis, but controls are added for the income, education, and family status of the imaginary worker being considered.

In Panels B and C, the outcome is an indicator for whether the difference in the number of stars a worker thought they themselves needed and the number of stars they thought the *different race same gender (1)* worker needed is positive or negative, respectively. Results are noisy, but in the manager sample, white women and racial minority men seem to anticipate positive and negative racial discrimination, respectively, when their manager knows demographic information. White women are 25 percent (statistically insignificant) less likely to think they need more stars than an equivalent Black woman and 80 percent more likely to think they need fewer stars than an equivalent Black woman, that is, they think they will be held to a lower standard. Racial minority men think they are 78 percent less likely to need fewer stars than an otherwise equivalent white man, that is, they think they will be held to a higher standard. White men also seem to anticipate positive racial discrimination (i.e. they anticipate being held to a lower standard), and are 65 percent less likely to think they need more stars than a Black man to be promoted when the manager can see their avatars, race, and gender than when the manager cannot, though this is statistically insignificant. In the algorithm sample, results are inconsistent. White women now are more likely to think they need more stars than otherwise equivalent Black women, and racial minority men are less likely to think they need more stars than otherwise equivalent white men. The sample sizes in the algorithm sample get quite small when splitting by racial group and gender in this way, however, which is why I do not focus on this race and gender heterogeneity in the algorithm sample in the main analysis of the paper.

In Panels D and E, the outcome is an indicator for whether the difference in the number of stars a worker thought they needed and the number of stars they thought the *same race different gender* worker needed is positive or negative, respectively. In the manager sample, the effect of being in the non-blind manager arm is only significantly different from zero for racial minority women, who seem to anticipate being held to a lower standard than men of the same race and equivalent other characteristics when managers observe avatars, race, and gender, though the effect of knowing that previously-promoted workers were mostly white men is the opposite in the algorithm sample.

Table I1: Effects on implicit measures of perceived discrimination in the manager sample of the promotion experiment

	Minority Women	White Women	Minority Men	White Men
Panel A: Number of additional stars I needed to be assigned harder task				
Non-blind manager, see avatars	0.061	-0.042	0.169*	-0.100
	(0.092)	(0.048)	(0.096)	(0.134)
Control mean	0.886	0.994	1.028	0.898
Panel B: I needed more stars than someone white (if racial minority) or black (if white)				
Non-blind manager, see avatars	-0.090	-0.046	0.019	-0.133
	(0.067)	(0.028)	(0.065)	(0.107)
Control mean	0.386	0.161	0.379	0.204
Panel C: I needed fewer stars than someone white (if racial minority) or black (if white)				
Non-blind manager, see avatars	-0.027	0.057**	-0.070**	0.000
	(0.038)	(0.022)	(0.032)	(0.076)
Control mean	0.079	0.068	0.090	0.041
Panel D: I needed more stars than a man (if woman) or a woman (if man)				
Non-blind manager, see avatars	-0.114*	0.010	-0.037	0.000
	(0.063)	(0.040)	(0.055)	(0.123)
Control mean	0.350	0.420	0.255	0.367
Panel E: I needed fewer stars than a man (if woman) or a woman (if man)				
Non-blind manager, see avatars	-0.007	-0.014	-0.029	0.062
	(0.041)	(0.017)	(0.040)	(0.069)
Control mean	0.107	0.054	0.103	0.020
N	265	728	289	94

Note: This table shows the effects of being in the non-blind manager arm on implicit measures of perceived discrimination relative to the demographic-blind manager arm in the promotion experiment. The implicit measures come from questions about how many additional "stars" (quiz-score quintiles) workers thought they and others would need to be promoted. In Panel A, the number of additional stars workers think they needed to be promoted; in Panels B and C, indicator for thinking they needed more or fewer stars than someone of another race, respectively; and the same but for gender in Panels D and E. Specifications are otherwise the same as in Panel A of 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 I2: Effects on implicit measures of perceived discrimination in the algorithm sample of the promotion experiment

	Minority Women	White Women	Minority Men	White Men
Panel A: Number of additional stars I needed to be assigned harder task				
See avatars	0.457**	0.045	0.036	-0.417
	(0.179)	(0.080)	(0.177)	(0.253)
Control mean	0.867	1.000	0.892	0.935
Panel B: I needed more stars than someone white (if racial minority) or black (if white)				
See avatars	0.094	0.085*	-0.191*	-0.094
	(0.112)	(0.051)	(0.098)	(0.177)
Control mean	0.407	0.138	0.268	0.261
Panel C: I needed fewer stars than someone white (if racial minority) or black (if white)				
See avatars	-0.151**	-0.021	0.018	0.010
	(0.060)	(0.038)	(0.068)	(0.069)
Control mean	0.111	0.057	0.171	0.022
Panel D: I needed more stars than a man (if woman) or a woman (if man)				
See avatars	0.076	-0.032	0.056	-0.292
	(0.101)	(0.060)	(0.098)	(0.182)
Control mean	0.272	0.258	0.195	0.348
Panel E: I needed fewer stars than a man (if woman) or a woman (if man)				
See avatars	-0.268***	-0.018	0.018	0.142**
	(0.066)	(0.024)	(0.084)	(0.062)
Control mean	0.173	0.038	0.146	0.043
N	148	313	150	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 implicit measures of perceived discrimination. The outcomes are the same as in Appendix Table I1. Specifications are otherwise the same as in Panel B of 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.