Does Information on Past Choices Affect Diversity in Hiring and Promotions? A Field Experiment with Hiring Managers at a Major Global Manufacturing Firm

Jeffrey Flory
(Claremont McKenna College)

Andreas Leibbrandt
(Monash University)

Olga Stoddard
(Brigham Young University)

We run a field experiment inside a large tech-oriented division at a Fortune 100 company to test a novel demand-side intervention in labor markets. The intervention was designed to reduce race and gender disparities, and lower potential biases based on these two candidate characteristics, in hiring and promotion in high-profile careers. The experiment, which includes over 5,000 applicants and over 125 managers, tests the effects of three different treatments that give hiring managers information on the gender and race composition of their past hires and promotions. Our subjects are actual managers, including very senior executives, who are treated while making genuine high-stakes choices over hiring into permanent positions at the firm, which also include quite senior levels and executive management. Our dataset is unique not only in the experimental treatments we test and the high economic stakes of the setting, but also in that we have data for applicant outcomes on all key stages in the hiring funnel – not just the initial screening for which applications pass the first stage, as is common in the literature. We find that, conditional on making it quite far through the process (not only being assessed as qualified but also being interviewed), nonwhites in the control group are significantly less likely to be offered an open position and also significantly less likely to be hired, compared to whites. We also find provocative evidence suggesting two of our treatments affect manager choices such that this race gap closes, making the probability of receiving a job offer and being hired comparable for white and nonwhite interviewees. Female candidates on the other hand are significantly *more* likely to be offered the open position and more likely to be hired in the control group. Our experimental treatments also appear to close the gender gap, making the likelihood that male and female interviewees are offered the job and hired comparable. While in both cases our treatments close the gap observed in the controls, in the first they do so to the benefit of traditionally underrepresented races/ethnicities in tech but in the second they do so to the disadvantage of a traditionally underrepresented gender in tech. These results highlight the power of showing managers overseeing hiring and promotions information on the race and gender of candidates they have chosen in the past to influence the race and gender characteristics among those they advance through the funnel and ultimately hire in the future.

1. Introduction

Important high-value sectors of the labor market continue to be marked by disproportionately low numbers of women and racial or ethnic minorities. Many employers have explicit goals to increase representation by women and minorities in the ranks of their employees, as well as the proportion they comprise at more senior levels, and they are heavily investing in initiatives to achieve these aims. However, the science on what works and how it depends on work setting, seniority level, timing during the recruiting process, and other circumstances, is still in many ways quite young. Firms are thus often left to try things on their own, with insufficient rigorous evidence to use as a guide. This paper, based on a field experiment inside a large tech-division at a fortune 100 company, examines a promising tool for addressing race and gender inequities in hires and promotions. We focus on a key part of the process through which open positions are filled, from the labor demand side – the hiring manager. By "holding up the mirror", to show managers data on the gender and race of their past hires and promotions, we uncover a potentially powerful approach to address gender inequities and discover important differences and limitations in how easily gender and race of new hires can be affected.

When it comes to race and ethnicity, policies aimed at raising the proportion of employees from underrepresented groups continue to lack solid empirical evidence on effective means of doing so. Companies and policy-makers are thus left with very little guidance and either do nothing or employ ad-hoc approaches that typically have no verifiable impacts and are also often expensive. Good intentions can backfire, as shown in work by Leibbrandt and List (2015), for example, where the inclusion of an "equal opportunity employer" statement for an employment position *lowered* application rates among ethnic minorities. Large resource investments on the part of firms or policy-makers may also lead to little or no effect. The identification of mechanisms to increase the number of workers in a given company from disproportionately underrepresented races or ethnicities thus represents a research area that is of critical policy relevance. It is also a research area that is wide open, with little rigorous empirical evidence, and many unanswered questions.

With respect to addressing gender inequalities in different employment sectors, more is known about certain features of the workplace that can skew the gender balance one way or the other. For example, there is a robust literature showing that workplaces and job tasks which involve individual competition have a strong tendency to push even high-skilled women away while they either attract men (including low-skilled men) or have a much weaker dissuasive impact on them. This can create sharp gender gaps wherein a majority of applicants are men, and a small percentage are women (and the skill distribution lower). This has been shown through a variety of experiments conducted in the lab as well as the field, in addition to survey-based studies (e.g. Niederle and Vesterlund, 2007; Niederle et al., 2013; Buser et al., 2014; Flory, Leibbrandt, and List, 2015; Flory et al., 2018).

Most prior work also tackles the question of gender and race/ethnicity imbalances in labor market outcomes from the labor-supply side. This includes work that identifies features of work settings that create gender or race disparities even among similarly skilled and qualified individuals (citations), as well as examining the efficacy of different approaches to encouraging greater interest and applications from women and ethnic minority groups (e.g. Flory, Leibbrandt, Rott, and Stoddard, 2021; Flory, Leibbrandt, Rott, and Stoddard, Forthcoming). In this study, we flip the perspective, tackling the question instead from the labor-demand side and the choices of actual hiring managers in deciding who to hire (or promote) into newly vacated positions within the company.

Some exceptions to the tendency to focus on the labor supply side include a vein of experiments that test for the existence of biases in hiring and worker evaluation, and examine the impacts of interventions to eliminate these biases, among student-subjects acting as workers and evaluators in the lab. For example, Bohnet et al. (2015) find significant reductions in gender biases in hiring, promotion, and job assignments through a “nudge” in which people are evaluated jointly rather than separately regarding their future performance. Flory, Leibbrandt, Stoddard (2023) examine the extent to which gender biases over expectations about future worker performance may grow or shrink when introducing on-the-job training and skill-development. However, studies in this vein have so far remained largely confined to the more controllable environment of the lab. When it comes to workplaces and companies that want to cultivate an environment supportive to the professional advancement and success of women and racial or ethnic minorities, much less is known about what works in the field, what doesn’t work, and what may even backfire and cause the *opposite* of intended effects.

This paper helps fill this gap in the research by implementing an experiment that tests a novel approach to influence the gender and race composition among external hires and internal promotions in a field setting with high stakes at a major corporation. Our subjects are managers in a large tech-oriented division of a Fortune 100 company, ranging from junior-level managers up through senior managers at the VP level. Treatments consist of showing managers statistics on the gender and race composition of their past hires and promotions in three different ways – in isolation, in comparison with peers, and when emphasizing their immediate supervisor can also see the information. Managers are treated while they are going about the normal process of hiring external candidates or promoting internal candidates into permanent jobs at the firm, which also range in seniority level from entry-level up through senior executive management. We are not aware of any other study that has successfully run a natural field experiment in employee recruiting and promotions that include such senior level employee subjects making such high stakes decisions about personnel.

The approach we test was intentionally designed to be cheap, easy to implement, and easy to scale up within the firm or across other firms if it had desirable effects. It was also designed so as to minimize risks of inciting or mobilizing any possible employee backlash or discontent with the new practice, by providing the personalized information treatment privately and at different times over a diffuse period. In addition, to maximize the impact of the information, treatments were carefully timed to be delivered only when a vacancy under a given manager's hiring authority opened, and to be delivered at key junctures during the decision-making process. To achieve this, and do so in a flexible way that accounts for unpredictability of when a vacancy will occur, we integrated treatment delivery with the firm's vacancy-tracking system, to be automatically sent to managers at certain points during their staffing decision. This therefore entailed minimal changes to the standard hiring procedures – the only change experienced from the manager's perspective was the inclusion of an extra line and a hyperlink (to the personalized information treatment) in a standard email each hiring manager receives periodically over the course of the vacancy-staffing process. Treatment delivery was thus nimble, and invisible to all but the manager-subject receiving the treatment.

The data the experiment generates is unique not only in the experimental treatments we test and the high economic stakes of the setting, but also in that we have data for applicant outcomes on all key stages in the hiring funnel – not just the initial screening of which applications pass through the first stage, as is common in the literature.

The main experiment sample consists of over 125 managers and over 5,000 applicants, and an augmented sample consists of over 200 managers and almost 10,000 applicants. We find interesting patterns both by gender and by race, but in opposite directions, as well as evidence that the treatments work to close gaps in the probability of moving through key stages in the hiring funnel. With respect to race or ethnicity, we find that conditional on making it quite far through the process (not only being assessed as qualified but also being interviewed), nonwhite applicants to positions hired into by managers in the control group are significantly less likely to be offered an opening as well as significantly less likely to be hired, compared to whites. After being interviewed, candidates of color are up to 21% (2.9 percentage points) less likely to be offered the job, and up to 20% (2.6 percentage points) less likely to be hired, compared to whites. We also find evidence suggesting two of our treatments affect manager choices such that this race gap closes. For example, in our expanded sample we find that showing a hiring manager the percentage of their own past hires and promotions that are women or people of color, alongside the same percentages for a peer group, increases the probability that a nonwhite candidate receives an offer from the manager by 2.3 percentage points ($p<0.05$). It also raises the probability that a nonwhite candidate is hired by 2.0 percentage points ($p<0.10$). We also find that just showing a manager their own statistics (without the comparison with peers) raises the probability a nonwhite will be hired by 1.8 percentage points ($p<0.10$).

For gender, the advantage in the control condition is instead experienced by the traditionally underrepresented group in the tech industry. Women had a 12%-14% (1.6 to 1.9 percentage point) higher probability than men of receiving a job offer after an interview, and a 7.4%-12.9% (0.9 to 1.7 percentage point) higher probability than men of being hired. Here again we find evidence that our treatments closed the gap. In the augmented sample, showing a manager the gender and race statistics of their own past hires and promotions alongside the same statistics among a peer group lowered the probability that a female candidate would receive an offer by 2.4 percentage points ($p<0.10$). Simply showing a manager their own statistics (without comparing with peers) lowered the probability a woman would be hired by 2.6 percentage points ($p<0.05$). Finally, showing a manager their own statistics, along with peer statistics and a message that their immediate supervisor could also see this information, lowered the probability a woman would be hired by 3.2 percentage points ($p<0.05$). In all three cases, the gender gap was not only erased but marginally to substantially reversed.

In sum, we find evidence of differential probabilities by gender and race in advancing through the hiring funnel in the status quo condition (i.e. absent the intervention). These differences occur at quite a late stage in the process (after applicants make it to the point of being interviewed), after a significant amount of vetting has occurred and presumably after all candidates have been judged as qualified. They also occur after a face-to-face interaction with the hiring manager. We also find evidence that manager choices about who to hire can be influenced by providing them with information about the gender and race of people they have selected in the past, such that existing race and gender gaps disappear, and even reverse.

The remainder of the paper is structured as follows. The next section explains our initial hypotheses, and the experiment we designed with the firm in order to test them. Section 3 examines the results from the field experiment. Section 4 concludes.

2. Hypotheses & Experiment Design

2.1 Hypotheses

{Discuss any literature suggesting that telling hiring managers the identity stats of who they hired in the past will change the identity characteristics of who they hire in the future, and what the potential mechanisms are.}

Hypothesis 1: Showing a hiring manager gender and race statistics of applicants they selected to fill openings in the past will push future their choices in a direction counter to past choices.

{Discuss any relevant literature on peer effects and social pressure.}

Hypothesis 2: Adding gender and race statistics of applicants selected to fill openings by a group of comparable peers, alongside gender and race information for a hiring manager's own past hires, will strengthen the effect.

{Discuss any relevant literature on boss pressure effects OR leadership signals that would cause us to have hypothesis 3.}

Hypothesis 3: Signaling to the hiring manager that their immediate supervisor may be assessing the gender and race characteristics of their hires, in addition to adding gender and race information of peer managers' hires, and their own past hires, will strengthen the effect even further.

2.2. Experiment Design – Overview

To test our hypotheses, we designed and integrated a field experiment within the normal hiring and promotions process at a large US manufacturing firm. The experiment took place in a large tech-oriented division (over 10,000 employees) at the company, a division that competes with other big tech employers for workers.

In addition to testing our main hypotheses, a few other motivations inform the particularities of our design. First, we wanted to test a method that would not only leverage the impacts of past demographic information on a manager's future hiring outcomes, but that would also be easy to implement and scale up within the firm, as well as across other firms, if it has desirable effects. Second, we wanted to minimize the potential for inciting opposition against the practice of employers giving managers this type of information (for example, if it were perceived as a directive to hire based on race or gender), or any backlash that might result from raising awareness around this new practice among many workers simultaneously. For example, if this type of manager-specific information were publicly shared, or even privately distributed but in a group meeting or coupled with a public announcement, the probability of inducing employee discussion on the topic and exacerbating discontent or articulating shared opposition may rise. Third, it was important to us that we minimize any other changes to the manager choice setting besides the delivery of the information itself in order to better isolate and attribute the cause of any effects as stemming from the information per se. Finally, to maximize the effectiveness of the information, we wanted it to be salient to the manager at a time when they are actually making hiring decisions, and for the timing of the treatment to be flexible and targeted enough to achieve this salience despite the unpredictability of when position openings would occur. The experiment was thus carefully designed to not only test our hypotheses, but also to: (i) be cheap and easy to implement (requiring few company resources); (ii) have no public visibility among employees and deliver treatments in a dispersed manner spread out over a long period – and only to the subset of managers who it could actually affect (those with an opening to fill); (iii) entail little disruption or change to the firm's standard hiring procedures (inserting a hyperlink to an information treatment in a standard email periodically sent to managers during the hiring process); and (iv) all while delivering a clear, personalized, private information treatment during a period when it was most likely to have an impact (when managers were actively filling a vacancy that happened to open).

One unique advantage of our design is that our subjects include managers of a broad variety of seniority levels (from the most junior managers up through VPs) who are hiring into permanent positions of employment that are also at a broad array of seniority levels – from entry level to quite senior executive management. We are not aware of any other study that has successfully run a natural field experiment in employee recruiting and promotions that include such senior level employee subjects making high stakes decisions about hiring into senior positions. Managers in the treatments are thus hiring into positions at a variety of different levels. This enables us to examine whether treatment effects differ by seniority of the position (if, for example, there is greater bias in hiring at more junior levels, or greater numbers of candidates of equal ability or less clear criteria for what constitutes the "best" candidate and thus more flexibility in selecting candidates).

The dataset generated by the experiment is a subset of the company's candidate-tracking database. This database consists of two types of individuals – applicants (who are tracked through a progression of stages that end when they are either cut from the pool, offered a job but decline, or are hired) and hiring managers (who oversee the progression of candidates through the hiring funnel). Applicants enter the candidate-tracking database by submitting an application for an opening at the firm. The positions include five different seniority levels at the company, ranging from entry-level up through and including senior executive management. There are several stages in the hiring funnel, beginning with…, and culminating in an offer and finally offer acceptance (hire). All applicants that submit an application to an opening during the experiment period become part of our data sample, and all managers that oversee hiring into any position within the division that becomes open during the experiment period are also part of our sample. Our dataset thus consists of the subset of the candidate-tracking database comprised by the data for these individuals during the experiment period.

The company's pre-existing standard procedure for interactions with managers charged with filling an open position is as follows. The manager first submits a request to fill an open position under their authority. This position may be internal or external , and is at one of five different seniority levels (from entry-level up to senior executive manager). The open position then appears in the vacancy-tracking system. Several days after the vacancy appears in the system, the hiring manager is sent an email with some basic information on what to expect next and to schedule a job launch call with a recruiter. (See Figure A for the email.) Roughly a week after this email is sent, the manager has a phone call with a recruiter to discuss the sourcing of applicants. Additional emails are sent to the manager at different stages during the process of filling the vacancy. These include: (i) when the vacancy is in the applicant-screening stage (see Figure B for email); (ii) once the slate of candidates has been sent to the manager (see Figure C); and (iii) when interviews are being conducted (see Figure D). The email sent to the manager when the vacancy is in the interview stage is sent twice (3 to 6 weeks apart) if the vacancy remains at the interviewing stage for more than 3-6 weeks.[1]

We integrate our experimental treatments into this pre-existing process to fill vacancies by tying them to the emails sent by HR to the hiring manager at each of the above stages. Working closely with HR, we had them insert a single sentence and a hyperlink near the top of each of these emails. The sentence was designed to maximize the likelihood that a manager would actually click the hyperlink, emphasizing to the manager that they may find the information personally important: "Click **here** {icon} to access new hiring statistics **YOU** may find interesting" (see Figures A through D). When a manager clicks on the hyperlink, they are first taken to a sign page where they enter the unique employee ID and password they typically use to access the company single sign-on system, and then proceed to a landing page (a standalone webpage) where they are shown their personalized treatment.

2.3. Experiment Design – Treatments

Table 1 summarizes the four treatments we implement in the experiment. Each treatment was designed to test one of our three main motivating hypotheses. They are layered such that each successive treatment includes the type of information provided in the prior treatment and builds on it. In the first treatment, the control condition, the landing page the manager is taken to contains a simple interesting fact about hiring at the company. This fact was designed to be at east mildly interesting (so that it fit the description about the hyperlink in the email) while having no plausible effect on the outcomes of interest: *"In 2017, {Company Name} accepted approximately {X} million*

[1] The standard process at the company is that a dedicated employee checks the status of all open vacancies once every 3 weeks and then sends whichever email corresponds to the category the vacancy is in at that time. Thus the vacancy may be at a given stage of being filled for anywhere between a day and 3 weeks before HR sends the corresponding email to the manager. The minimum amount of time a vacancy could be at the interview stage and cause 2 emails to be sent is just over 3 weeks, and any vacancy that is at the interview stage for more than 6 weeks will receive two emails.

applications and posted more than {Y},000 jobs in over {Z} countries." See Figure E for what the landing page looked like under the control condition.

Table 1. Experiment design

Hypothesis	Message type	Treatment	Description
H1	*Control*	T0	General fact about hiring.
H1	*Personal*	T1	T0 + Personalized Past Hiring and Promotions Gender & Race Statistics (PHP)
H2	*Personal/Peer*	T2	T1 + Peer Group Past Hiring and Promotions Gender & Race Statistics
H3	*Personal/Peer + Supervisor*	T3	T2 + Statement that immediate supervisor also sees this information

Manager-subjects assigned to the second treatment group (first experimental treatment, T1) see the same general fact as in T0 when they click on the hyperlink, and also see information on their own record of gender and race statistics in their past hires and promotions (PHP) up until about 4 months before the experiment launch. To make the information clear and salient, it is given in two different bar graphs – one for gender and one for race – with the exact percentage of their hires and promotions that are women or that are nonwhite written in bold at the top of the corresponding bar. At the top of the screen (below the general fact) in large letters is the title "Hiring Track Record as of April 2018". The title for each graph reads "Gender Hires & Promotions" and "Racially Diverse Hires & Promotions." To help ensure the manager knows this is their own record, "your" is in all caps. The statement below the gender graph reads: "% Women in YOUR hires/promotions." The statement below the race graph reads: "%EEO diversity in YOUR hires/promotions" See Figure F for an example of this treatment for a manager whose hires and promotions were 32% women and all white.

Managers in the third treatment group (second experimental treatment, T2) who click the hyperlink see the same fact as in T0, see their own PHP statistics (just as in T1), and also see the percentage of past hires and promotions that were women and nonwhite in a peer group. The managers selected to create the peer group were not random or representative – they were selected so as to push the percentage of women and racial minorities somewhat higher than the actual average. Besides the information on peers, everything else about the landing page is the same as in T1. To facilitate easy comparison, the peer group percentages are shown as a separate column (different color) next to the manager's own column, for both gender and race information. The statement below the gender (race) graph for the peer group reads: "% Women (EEO diversity) in your PEERS' hires/promotions." See Figure G for an example of this treatment for a manager whose hires and promotions were 80% women and all white.

In the final treatment group (T3), managers who clicked on the link were shown the same information as in T2, with an additional statement highlighted in bold orange at the bottom that indicated their own manager can see this information. This statement read: "This information is available to your immediate manager." Aside from this statement, the landing page looked exactly the same as in T2. See Figure H for an example.

Only managers who had at least 5 years (??) of experience hiring/promoting within the company were included in the experiment, since we decided that PHP information for those with little hiring and promotions experience did not make sense. Managers were randomized into treatment groups, and their landing page bar graphs generated, before the start of the experiment. We assigned over 1,000 managers in the division to a treatment. A little under half of these had an opening occur under their authority over the course of the experiment.

3. Results

3.1. Summary Statistics

The experiment ran from August 1, 2018 to December 31, 2019. Over the course of these 17 months, a total of 478 managers had one or more positions become open and were thus sent the emails with links to their treatment landing page. Among these managers, 14.6% are women, and 19.8% are non-white. These managers made 1,551 job offers, 1,427 (92%) of which were accepted by the applicant. However, gender and race data are missing for a substantial number of applicants. Restricting to cases where gender (race) is known, 967 (956) offers were made and 865 (854) were accepted, and restricting to cases where gender and race are both known (gender OR race is known) XX (YY) offers were made and ZZ (XX) were accepted.

There were a total of 27,601 job applications to the roughly 1,500 positions that became open during the experiment, 5.6% of which received offers and 5.2% of which became hires.

Among the 23,263 applicants (84%) with gender data, 22.1% are women. Among the 22,965 (83%) with race data, 41.7% are nonwhite: 18.3% Asian, 12.2% Black, 7.5% Hispanic, 3.1% mixed and other.

Since managers are not forced to look at their PHP landing page statistics, but rather must click on the link in order to see it, the design is an intent to treat. Of the 478 managers assigned to treatments who had a vacancy and were thus invited to click on the link, 129 managers (27%) opened the link at least once. As expected (since treatment was randomly assigned before the experiment), this group is pretty evenly spread across the treatments: 34 in T0, 30 in T1, 31 in T2, and 34 in T3. In total, among job offers for which both gender and race data for the candidates offered the position are complete, 192 offers were extended from these treated managers: 47 from managers in T0, 40 from those in T1, 44 from those in T2, and 51 from those in T3.

3.2. Experimental Results

We examine the results in a few different ways. The first approach examines outcomes at the applicant level, estimating the average effect of each manager treatment on the likelihood that a candidate of a given gender or race makes it through each key stage in the hiring funnel. There are five key stages in the hiring process – application reviewed, applicant is qualified, manager is interested in applicant, applicant receives job offer, applicant accepts job offer. We examine treatment effects both through non-parametric analyses and linear regressions. For each stage, the sample consists of the applicants who have made it past the previous stage – i.e. we examine treatment impacts on making it through each stage conditional on making it to that stage. The main regression equation for estimating effects on candidates making it through each funnel stage in this analysis is:

$$\begin{aligned} y_{im} = {} & \beta_0 + \beta_1 \mathrm{T1}_m + \beta_2 \mathrm{T2}_m + \beta_3 \mathrm{T3}_m + \\ & \gamma_1 F_{im} + \gamma_2 F_{im} \mathrm{T1}_m + \gamma_3 F_{im} \mathrm{T2}_m + \gamma_4 F_{im} \mathrm{T3}_m + \\ & \theta_1 R_{im} + \theta_2 R_{im} \mathrm{T1}_m + \theta_3 R_{im} \mathrm{T2}_m + \theta_4 R_{im} \mathrm{T3}_m + \epsilon_{im} \qquad (1) \end{aligned}$$

where $y_{im} = \{0,1\}$ is an indicator for whether the i^{th} applicant under the m^{th} manager made it through the given stage in the hiring funnel. T1-T3 are treatment indicators for whether the m^{th} manager experienced the given treatment. F is an indicator for female, and R is a race indicator equal to one if the applicant is nonwhite. The omitted category is thus white males. Standard errors are clustered at the manager level, the level of random assignment of the treatment {OR:} The specification is a random effects GLS model, where the group variable is the manager. This analysis tests for effects by gender and race separately.

We restrict our analysis to those who were actually treated – i.e. the compliers. However, in some analyses we relax this restriction for those in the control group under the assumption that the control message factoid had no impact on manager behavior (which we test) and that compliers are non-compliers in this setting do not systematically differ in ways that affect the response variables of interest (which we discuss and also test). Further below, we discuss what the estimated effects for compliers can also tell us about likely effects for the noncomplying hiring managers.

We first focus on two key outcomes – receiving a job offer, and getting hired – and we include all the data from the experiment period (August 14, 2018 to August 14, 2019). Table 2 shows estimates from regression equation 1. Columns 1 and 3 restrict the sample to compliers – i.e. managers who clicked on the link and were shown the landing page in their randomly assigned treatment group. As the constant term in column 1 shows, among white males who were interviewed (i.e. advanced through the third stage in the hiring funnel), 13.2% were given a job offer. As the coefficient for *Female* shows, women had a 1.6 percentage point (12%) higher probability than men of receiving a job offer after an interview, with a total of 14.8% of white women receiving an offer, though this increased probability of women compared to men is not statistically significant at conventional levels (p=…). Turning to the coefficient for *Nonwhite*, we see that applicants of color had a 2.0 percentage point (15%) lower probability of receiving an offer after an interview, with a total of 11.2% of those interviewed receiving an offer, and this lower probability compared to whites is statistically significant ($p<.10$). Turning to the experimental treatments, we find no significant estimates for treatment effects in this specification. However we note that the estimated magnitudes for the interaction terms *T1×Female*, *T2×Female*, and *T3×Female* are all the opposite sign and similar magnitude as the estimate for *Female*, suggesting that these treatments, to the extent that they have any effect, actually neutralize any advantage that women experienced in making it through this stage (receiving an offer after being interviewed).

Column 3 shows results for the regression of whether or not an applicant was ultimately hired (was offered the position and accepted it), conditional on being interviewed (i.e. making it through the third stage). The constant term shows that 11.6% of white males who were interviewed were ultimately hired. This is about 1.5 percentage points lower than the 13.2% of white male interviewees who were offered a job, indicating a rejection rate of about 13% among white men (i.e. 13% of white males offered a position did not accept it). Turning to the coefficient for *Female*, we see it remains non-significant and that its magnitude falls by about a half. About 12.5% of white women were ultimately hired, indicating a rejection rate of about 18% among white women. The coefficient for *Nonwhite* on the other hand grows in magnitude and becomes more significant: applicants of color had a 2.2 percentage point (18.9%, $p<0.05$) lower probability of being hired after being interviewed, compared to whites. For example, only 9.4% of nonwhite males were ultimately hired after an interview compared to 11.6% of white males. The rejection rate among nonwhites is also larger: among nonwhite males the rejection rate was about 19% (compared to 13% among white men). This may suggest inferior offers (e.g. lower salary or other benefits) to nonwhites (relative to outside options) compared to whites. As for estimated treatment effects, we again see that none of the estimates are statistically significant in this specification for impacts on probability that a candidate is hired.

Columns 2 and 4 repeat the above two regressions but augment the sample by expanding the control group to include noncompliers. That is, for T0, rather than including only the applicants to positions where the hiring manager clicked the link and saw the landing page, we also include in the analysis applicants to positions where the hiring manager was randomly assigned to T0 but did not click the link. Recall that the landing page for T0 had an innocuous factoid unlikely to affect hiring managers' choices about who to advance through each stage in the hiring funnel.[2] Indeed, when we test to see if the choices of managers randomly assigned to T0 differ between compliers (clicked link and saw their PHP) and noncompliers (did not click link and so were not shown their PHP), there are no significant differences in the outcome variables for the last two stages (offers and hires), overall or by gender or nonwhite status. Thus, as long as compliers and noncompliers are not systematically different in ways that impact the outcome variables or treatment effects on the outcome variables, the noncompliers in T0 are a valid comparison group for the compliers in T1-T3. Pooling the compliers and noncompliers in T0 into a pooled control

[2] T0 has no information about gender or race, only a simple fact about quantity of applicants, jobs, and countries where the company hired. There is no clear reason that this would alter the behavior of managers with respect to the gender or race of applicants they advance through each stage.

group substantially expands the sample – by about 70%, from 129 managers to 217 managers, and from 5,679 applicants to 9,815 applicants.

In this pooled control specification, we see in Columns 2 and 4 that the coefficient estimates are very similar to those in the corresponding regressions in Columns 1 and 3 (especially for estimates measured with any degree of precision). We also see that the precision of estimates in the expanded sample is substantially higher, with more of the estimates significant at conventional levels – including several of those for estimated treatment effects. For example, the elevated probability that women experience in both receiving an offer (conditional on being interviewed) and in being hired (conditional on being interviewed) are now highly significant – but only in the control group. As column 2 shows, women are 1.9 percentage points ($p<0.01$), or 14%, more likely than men to be offered the opening after an interview when managers overseeing the hiring process are in the control group. However all three experimental treatments appear to nullify this advantage that women enjoy over men in terms of probability of job offer: the coefficients for the interaction terms *T1×Female*, *T2×Female*, and *T3×Female* are all roughly similar in magnitude to the coefficient for *Female* but with the opposite sign, driving the impact of gender on being offered the vacancy back down towards zero. One of the coefficients (*T2×Female*) is now significant ($p<0.10$): Treatment T2 (showing the manager their own PHP alongside their peers' PHP statistics) reduced the likelihood that women would be given an offer by 2.4 percentage points (17%). For white women, for example, this drives their probability of receiving an offer down to an estimated 13.1% and making it statistically indistinguishable from that for white men (a 13.5% probability).

We see quite similar patterns for gender when it comes to ultimately being hired. Column 4 shows that women are 1.7 percentage points (12.9%) more likely than men to be hired (receive and accept a job offer) after receiving an interview, and that this difference is highly significant ($p<0.01$). Once again, however, the experimental treatments eliminate this gender gap that favors women, causing men and women who make it through the interview stage to be hired with similar probability. Coefficients for the interaction terms *T1×Female*, *T2×Female*, and *T3×Female* show that T1 lowers the probability women are hired by an estimated 2.6 percentage points compared to T0 ($p<0.05$), T2 lowers it by an estimated 2.0 percentage points (not significant), and T3 lowers it an estimated 3.2 percentage points ($p<0.05$). All of these estimates are large enough to counteract – and even slightly reverse – the gender gap in probability of being hired we see in the control group, which favors women.

Turning to race, we see a similar pattern – but in this case the status quo advantages traditionally overrepresented demographics in STEM, and then treatments induce manager choices more favorable to underrepresented groups. Column 2, for example, shows that nonwhites are a full 2.9 percentage points (21.4%) less likely than whites to be offered the vacancy conditional on being interviewed. In addition we see in the coefficient estimates for *T1×Nonwhite*, *T2×Nonwhite*, and *T3×Nonwhite* that the three experimental treatments tend to push manager choices back against this race gap that favors whites. All three coefficients are positive, though only one is significant. In particular, treatment T2 (showing the manager their own PHP alongside their peers' PHP statistics), raised the probability nonwhites would be offered the job by an estimated 2.3 percentage points ($p<0.05$), closing the race gap in job offers.

Similarly, when examining the racial characteristics of applicants who are actually hired, we once again find a large race gap is the status quo, and that our experimental treatments tend to close this gap. Results in column 4 show that being nonwhite lowers an applicant's probability of being hired into the position by an estimated 2.6 percentage points ($p<0.01$), which represents a 20.2% fall between white and nonwhite men and a 17.9% fall among women. Once again, however, the coefficients on the treatment-race interaction terms show the experimental treatments tend to close this race gap. This time, two are significant at conventional levels: treatment T1 (showing a manager their PHP) raises the probability nonwhites will be hired by 1.8 percentage points ($p<0.10$), and treatment T2 (own PHP alongside their peers' PHP statistics) raises the probability nonwhites are hired by 2.0 percentage points ($p<0.05$) – both treatments pushing the probability nonwhites are hired back towards parity with whites.

The results for impacts on women suggest a reduction in bias, but in a surprising way. Whereas in the control group women enjoyed a *higher* probability of receiving a job offer and ultimately being hired compared to men, the treatments tended to *decrease* this probability, putting it back on parity with that of men.

4. Conclusion

This study sheds light on mechanisms to eliminate inequalities with respect to gender and race in labor market outcomes and employment – with an emphasis on professional ("white collar") and management-level employment. We run a natural field experiment with manager subjects in a large tech division of a Fortune 100 company to examine impacts on high-stakes staffing decisions of a novel intervention designed to close gender and race gaps in hiring and promotions. The data we generate is unique not only in that it is a demand-side intervention including senior managers and high-level vacancies in a STEM-oriented division of a major firm, but also in that it includes observations on the full range of outcomes throughout the hiring funnel – from application submission to being hired.

The data from our field experiment highlight two main conclusions. The first is that, absent the intervention, there were important differences by gender and race in the probability of advancing through the hiring funnel. These differences exist at quite a late stage in the process (after applicants make it to the point of being interviewed), after a significant amount of vetting has occurred and presumably after all candidates have been judged as qualified. They also occur after a face-to-face interaction with the hiring manager. Interestingly, while both women and racial minorities are often underrepresented in STEM settings in the workplace, the gender gap in probability to receive an offer or be hired actually favors women, whereas the race gap favors whites.

We also find evidence that manager choices about who to hire can be influenced by providing them with information about the gender and race of people they (and their peers) have selected in the past, such that existing race and gender gaps disappear, and even reverse. Both initial gaps were erased by one or more of our treatments for one or more of the outcomes.

Methodologically, the study also demonstrates a new way to implement small changes in hiring practices at large employers in a manner that is nimble, flexible, cheap, and easy to scale. By linking the delivery of an information intervention for decision-makers in the hiring process with an organization's vacancy tracking system in a manner that is automatic, managers can receive the information privately, at a time when it is likely to have the most impact, and in a way that creates little disruption to the pre-existing process and minimizes the likelihood for employee backlash or discontent.

References

Bohnet, Iris, Alexandra Van Geen, and Max Bazerman. "When performance trumps gender bias: Joint vs. separate evaluation." *Management Science* 62.5 (2016): 1225-1234.

Buser, T., M. Niederle, and H. Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *Quarterly Journal of Economics* 129(3):1409–47.

Flory, J.A., U. Gneezy, K.L. Leonard, and J.A. List. 2018. "Gender, Age, and Competition: A Disappearing Gap?" *Journal of Economic Behavior and Organization* 150:256–76.

Flory, J.A., A. Leibbrandt, and J.A. List. 2014. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *Review of Economic Studies* 82(1):122–55.

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* 56, no. 1 (2021): 73-92.

Flory, Jeffrey A., Andreas Leibbrandt, Christina Rott, and Olga Stoddard. "Leader signals and the power of growth mindset: A natural field experiment in workplace diversity. *Management Science*, forthcoming.

Leibbrandt, A., and J. List. 2018. "When Equal Employment Opportunity Statements Backfire: Evidence from a Natural Field Experiment on Job-Entry Decisions." NBERWorking Paper 25035. Cambridge, MA: NBER.

Niederle, M., and L. Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics* 122(3):1067–1101.

Niederle, M., C. Segal, and L. Vesterlund. 2013. "How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." *Management Science* 59(1):1–16.

Tables

Table 2

	(1)	(2)	(3)	(4)
	offer	offer	acceptance/hires	acceptance/hires
sample	only managers who have seen treatment	all managers in T0 , T1-3: only those who saw treatment	only managers who have seen treatment	all managers in T0 , T1-3: only those who saw treatment
T1	-0.0194	-0.0229	-0.0084	-0.0193
	(0.7629)	(0.6732)	(0.8919)	(0.7187)
T2	0.0011	-0.0024	0.0166	0.0060
	(0.9866)	(0.9650)	(0.7865)	(0.9097)
T3	0.0231	0.0196	0.0317	0.0207
	(0.7113)	(0.7047)	(0.5973)	(0.6857)
Female	0.0160	0.0187***	0.0086	0.0165***
	(0.2147)	(0.0031)	(0.4817)	(0.0054)
T1 x Female	-0.0186	-0.0212	-0.0177	-0.0256**
	(0.2884)	(0.1231)	(0.2838)	(0.0478)
T2 x Female	-0.0210	-0.0236*	-0.0123	-0.0202
	(0.2275)	(0.0823)	(0.4528)	(0.1142)
T3 x Female	-0.0143	-0.0169	-0.0243	-0.0323**
	(0.4513)	(0.2812)	(0.1741)	(0.0289)
Non-white	-0.0195*	-0.0290***	-0.0220**	-0.0259***
	(0.0513)	(0.0000)	(0.0192)	(0.0000)
T1 x Non-white	0.0041	0.0137	0.0141	0.0180*
	(0.7564)	(0.1854)	(0.2573)	(0.0636)
T2 x Non-white	0.0131	0.0226**	0.0156	0.0196**
	(0.3219)	(0.0273)	(0.2078)	(0.0426)
T3 x Non-white	-0.0088	0.0008	-0.0015	0.0024
	(0.5645)	(0.9503)	(0.9140)	(0.8462)
Constant	0.1317***	0.1353***	0.1164***	0.1283***
	(0.0029)	(0.0000)	(0.0061)	(0.0000)
N	5679	9815	5679	9815
Managers (grouping var)	129	217	129	217

Notes: Random-effects GLS model. Group variable: Manager

Appendices

Figure A: FIRST EMAIL – JOB LAUNCH*

Email Subject: "Your Vacancy Was Approved…Here Are Your Next Steps!"

{Company Logo}

Talent Acquisition

{Division Name} CoE

{Quote from the President and CEO of the company division about the importance of human capital, leadership, innovation, and ideas to drive company success}

{Name}, President and CEO, {Company Division}

Click here to access new hiring statistics YOU may find interesting.

Your vacancy has been received by the {Company Division} Talent Acquisition COE. It will be reviewed to ensure it meets compliance standards then posted to the appropriate gateways.

What Happens Next?
Your assigned recruiter will reach out to schedule a call to officially kick off the recruiting process. Your participation on this call is critical!

* This is an example of what the actual email looked like. Several features have been altered to keep the company anonymous.

Figure B: SECOND EMAIL – SCREENING STAGE*

Email Subject: "You Are a Talent Ambassador!"

{Company Logo}

Talent Acquisition

{Division Name} CoE

{Quote from the President and CEO of the company division about the importance of human capital, leadership, innovation, and ideas to drive company success}

{Name}, President and CEO, {Company Division}

Click here to access new hiring statistics YOU may find interesting.

Your recruiter is actively sourcing qualified candidates for your open vacancy. You can partner in this process by encouraging referrals from your team and sharing your open position with your personal and professional networks.

Consider connecting with your recruiter to understand how you can be a Talent Ambassador!

* This is an example of what the actual email looked like. Several features have been altered to keep the company anonymous.

Figure C: THIRD EMAIL – SLATE STAGE*

Email Subject: "Your Role in Assessing Talent"

{Company Logo}

Talent Acquisition
{Division Name} CoE

{Quote from the President and CEO of the company division about the importance of human capital, leadership, innovation, and ideas to drive company success}

{Name}, President and CEO, {Company Division}

Click here to access new hiring statistics YOU may find interesting.

By now, you should have received profiles of candidates to review for your open vacancy. Your diligence in assessing skills and capabilities will help to ensure that we identify the very best talent to move forward in the selection process.

* This is an example of what the actual email looked like. Several features have been altered to keep the company anonymous.

Figure D: FOURTH EMAIL – INTERVIEW STAGE*

Email Subject: "The Interview Process…First Impressions Matter!"

{Company Logo}

Talent Acquisition

{Division Name} CoE

{Quote from the President and CEO of the company division about the importance of human capital, leadership, innovation, and ideas to drive company success}

{Name}, President and CEO, {Company Division}

Click here to access new hiring statistics YOU may find interesting.

As a Hiring Manager, you serve as ambassador of our {Company Name} brand. The onsite experience helps the candidate understand what day-to-day life will be like working on your team. Consider the logistics of the interview day and be sure to incorporate touches that will leave a favorable first impression!

* This is an example of what the actual email looked like. Several features have been altered to keep the company anonymous.

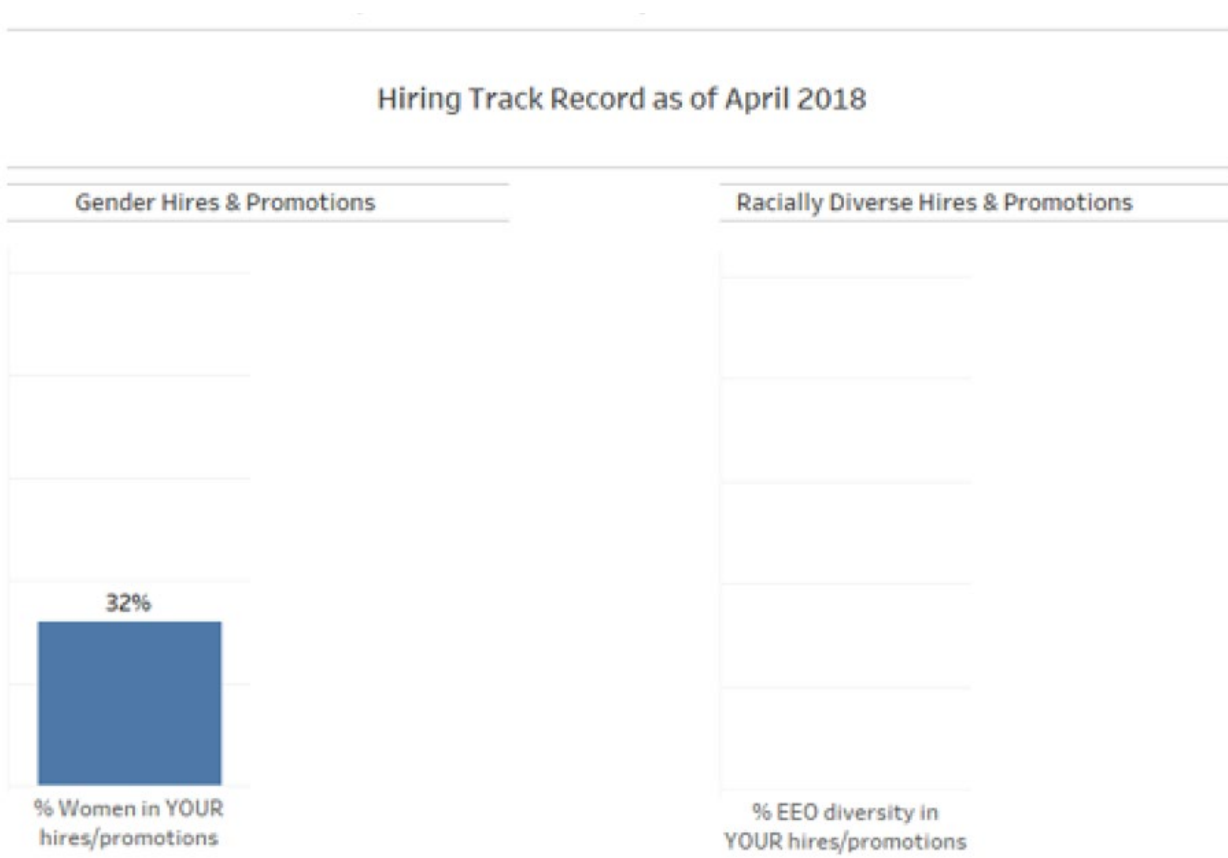
Figure E: Treatment T0 (General Fact)

In 2017 {Company Name} accepted approximately {X} million applications and posted more than {Y},000 jobs in over {Z} countries.

Please submit any feedback to: mailto: {Email of a dedicated employee at the company.}

Figure F: Treatment T1 (T1 + Personalized PHP Stats)

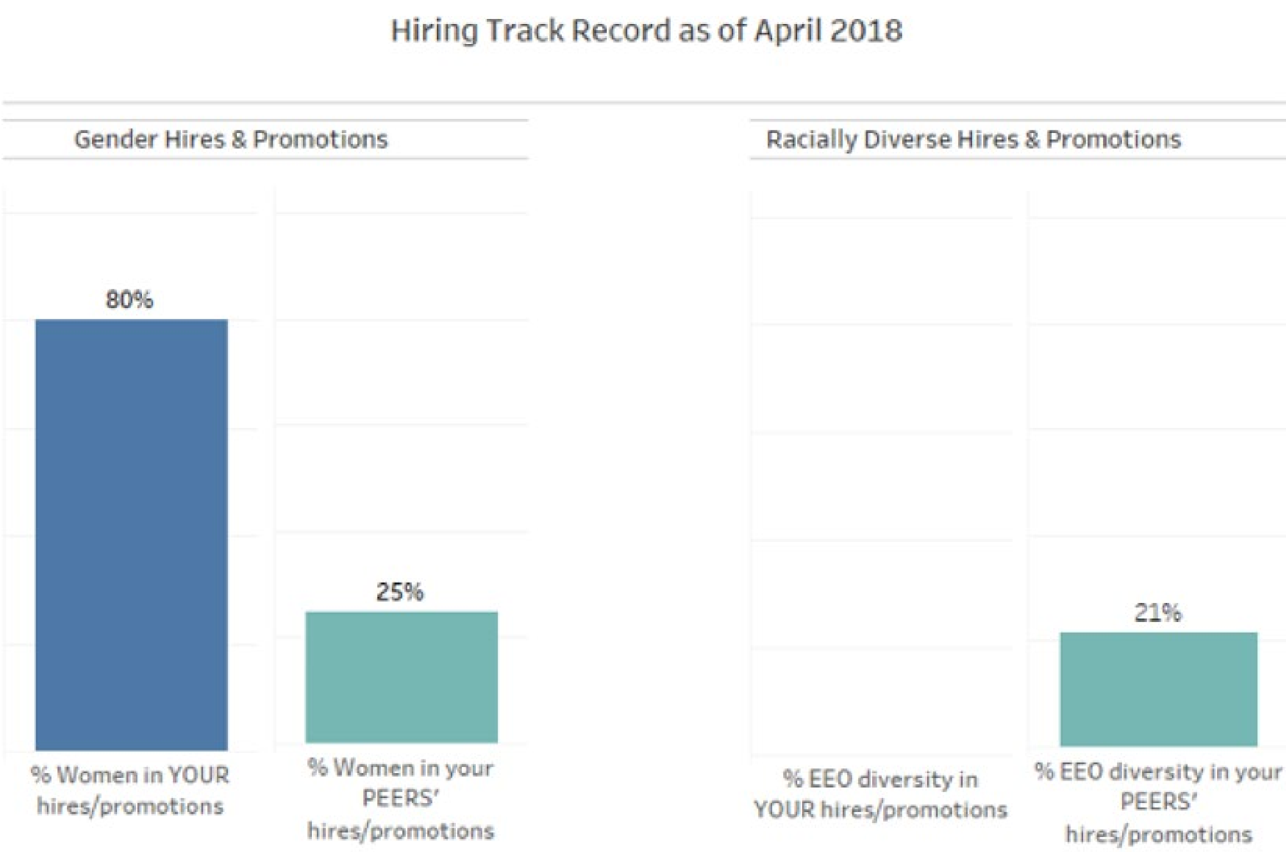
In 2017 {Company Name} accepted approximately {X} million applications and posted more than {Y},000 jobs in over {Z} countries.



Please submit any feedback to: [mailto: {Email of a dedicated employee at the company.}]

Figure G: Treatment T2 (T1 + Peer Group PHP Stats)

In 2017 {Company Name} accepted approximately {X} million applications and posted more than {Y},000 jobs in over {Z} countries.



Please submit any feedback to: mailto: {Email of a dedicated employee at the company.}

Figure G: Treatment T2 (T1 + Peer Group PHP Stats)