Discrimination in Hiring: Evidence from Retail Sales*

Alan Benson[†], Simon Board[‡], and Moritz Meyer-ter-Vehn[§]

April 2, 2023

Abstract

We propose a simple model of racial bias in hiring that encompasses three major theories: taste-based discrimination, screening discrimination, and complementary production. We derive a test that can distinguish these theories based on the mean and variance of workers' productivity under managers of different pairs of races. We apply this test to study discrimination at a major U.S. retailer using data from 48,755 newly-hired commission-based salespeople. White, black and Hispanic managers within the same store are significantly more likely to hire workers of their own race, consistent with all three theories. For black-Hispanic pairs, productivity variance is lower for same-race pairs than cross-race pairs, implying that screening discrimination dominates. For white-Hispanic pairs, mean productivity is higher for same-race pairs, indicating a combination of screening discrimination and complementary production. For white-black pairs, biased hiring implies the presence of discrimination, but productivity results suggest the effects of the three forms of discrimination offset one another.

1 Introduction

More than fifty years after the U.S. prohibited employers from using race as a factor in employment, the persistence of racialized labor market outcomes—including segregation and pay gaps—remains among the most striking features of the labor market. Controlling for education and experience, black and Hispanic workers have significantly lower employment rates and wages than white workers, especially for less-educated workers (Altonji and Blank, 1999; Ritter and Taylor, 2011; Lang and

*Keywords: discrimination, screening, hiring. JEL codes: J71, M51. We thank Aislinn Bohren, Adriana Lleras-Muney, Mandy Pallais, Andres Santos and seminar participants at the AEAs, Minnesota, MIT, NBER, Penn, and UCLA. We also thank the W.E. Upjohn Institute for Employment Research for support.

†University of Minnesota, https://carlsonschool.umn.edu/faculty/alan-benson

‡UCLA, http://www.econ.ucla.edu/sboard/

§UCLA, http://www.econ.ucla.edu/mtv/

Lehmann, 2012). Understanding the causes of racial disparities is important for policymakers and managers who wish to promote equality and enhance productivity.

In this paper, we study how workers' outcomes depend on the race of their hiring manager, an idea known as *relative* employment discrimination. We seek to understand why such discrimination arises by studying three major theories. Taste-based discrimination (Becker, 1957) is the simplest, and perhaps the presumptive, theory of discrimination. If employers have an intrinsic taste for hiring same-race workers then they lower the hiring bar for their own race, leading to labor market segregation and wage disparities (e.g. Black, 1995). Screening discrimination (Cornell and Welch, 1996), though less prominent in the literature, poses an alternative theory of racialized labor market outcomes. This model adapts Phelps's (1972) classic model of statistical discrimination to allow employers to be better at screening same-race applicants.[1] The third theory of discrimination is complementary production (e.g. Lang, 1986) in which workers are more productive when working with managers of the same race, leading to higher rates of same-race hiring.[2] All three theories are consistent with a wide variety of empirical facts (e.g. labor market segregation, pay gaps, and differences in employment rates), but call for different remedies. For example, racial quotas raise productivity under taste-discrimination, as biased managers hire more qualified minorities, but lower it under screening discrimination, as uninformed managers hire mismatched minorities. Conversely, algorithmic hiring tools that provide managers information about candidates improve diversity and productivity under screening discrimination, but not under taste-based discrimination.

This paper proposes a common framework to distinguish and test for these three theories of discrimination. In the model, managers observe a noisy signal of applicants' productivity and hire those whose expected productivity lies above a threshold. With taste-based discrimination, managers adopt lower thresholds when evaluating same-race applicants. With screening discrimination, managers observe a more precise signal when evaluating same-race applicants. And with complementary production, managers obtain higher output when employing a same-race worker.

All three theories predict that managers hire more same-race applicants, although for different reasons. With taste-based discrimination, managers lower their standards for same-race candidates; with screening discrimination, managers receive stronger signals of ability from same-race candidates; with complementary production, managers obtain higher output for a given worker if they share the manager's race. All three theories also predict same-race workers have lower turnover. With taste-based discrimination, managers treat same-race workers more leniently; with screening discrimination, managers make fewer mistakes in hiring same-race workers; with complementary production, same-race workers are less likely to be marginal hires and at risk of being fired. Thus

[1] Statistical discrimination captures the idea that firms have more accurate signals of majority applicants. Screening discrimination captures the idea that managers obtain more accurate signals from same-race applicants.

[2] Such "complementarity production" could be because of closer cultural fit or because of prejudice. Complementarity means the most productive assignment is to match workers with same-race managers; it contrasts with "favoritism", like assigning better shifts to same-race workers, that is zero sum.

Table 1: **Summary of theoretical results**

	Is same-race or cross-race higher?			
	Hiring	Prod. mean	Prod. variance	Turnover
Taste-based discrimination	Same-race	Cross-race	Same-race	Cross-race
Screening discrimination	Same-race	Same-race	Cross-race	Cross-race
Complementary production	Same-race	Same-race	Same-race	Cross-race

one cannot separate between the three theories using hiring or turnover data.

In contrast, the three theories make different predictions regarding the relative productivity of same-race and cross-race hires. Under taste-based discrimination, managers lower the hiring threshold for same-race workers, meaning that the productivity of same-race hires has a lower mean and higher variance. Under screening discrimination, managers can screen same-race workers with greater accuracy, so the productivity distribution of same-race hires has a higher mean and lower variance. And under complementary production, productivity of same-race workers shifts up, so the productivity distribution of same-race hires has a higher mean and higher variance. These theoretical results are summarized in Table 1.

Using these predictions, we derive a test for discrimination in a model that nests the three theories and allows for a small amount of manager and worker heterogeneity (e.g. different races of workers may have different means and different races of managers may have different hiring thresholds). Our test is based on the sub/supermodularity of the productivity mean and variance.[3] If mean productivity is submodular in manager-worker race, then there exists taste-based discrimination. If log productivity variance is submodular, then there exists screening discrimination. And if mean productivity and log productivity variance are both supermodular, then there exists complementary production. The first two results follow immediately from Table 1 but the third is more subtle: No combination of the other forms of discrimination can produce both types of supermodularity simultaneously.

Our test is conservative in that it avoids false positives but can lead to false negatives. First, it may be that all three forces are present, but we identify none because they offset one another; indeed, we can identify at most two forces from the same data set. Second, our test can detect relative discrimination but cannot tells us if all managers are biased against a particular race of worker.

We apply our model to longitudinal administrative data from a large national retailer from 2009

[3]If $\bar{y}_{ij}$ is the mean productivity of race-i workers hired by race-j managers, then $\bar{y}$ is *supermodular* if $\bar{y}_{ii} + \bar{y}_{jj} \geq \bar{y}_{ij} + \bar{y}_{ji}$, and *submodular* if the inequality is reversed. Similarly, if v_{ij} is the productivity variance of race-i workers hired by race-j managers, then v is *log-supermodular* if $\log v_{ii} + \log v_{jj} \geq \log v_{ij} + \log v_{ji}$, and *log-submodular* if the inequality is reversed.

to 2015. Each store typically employs several hiring managers of different races, allowing us to examine the difference in their behavior after controlling for store, department and month fixed effects. The data include the races of 48,755 white, black, and Hispanic newly-hired salespeople and the races of their 7,892 hiring managers at 997 store locations over 74 months.[4] The data further include sales versus targets, the performance measure used to determine commissions.

First, we show that the probability that a hire is a given race is 3.6pp higher when the manager shares the same race, after controlling for fixed effects. This gap is positive and statistically significant for all same-race pairs, and represents a particularly large increase for minority workers since they account for a much smaller percentage of hires. For example, a Hispanic worker is 25.5% more likely to be hired if the hiring manager is also Hispanic. These results are consistent with all three theories.

Next, we use the productivity data to shed light on which of the three theories causes this discrimination in hiring. The cleanest evidence comes from black-Hispanic pairs, where the mean gap between same-race and cross-race productivity is 2.1pp while the same-race variance is 82% of the cross-race variance. We reject the null that there is no screening discrimination at the 5% level. Indeed, at the bottom of the distribution, the gap between same-race and cross-race productivity is 10pp, which suggest managers eliminate low-productivity same-race workers. For white-Hispanic pairs the mean gap between same-race and cross-race productivity is 1.4pp while the same-race variance is 96% of the cross-race variance, and we reject the null that there is neither screening discrimination nor complementary production. For white-black pairs the mean gap between same-race and cross-race productivity is 0.7pp while the same-race variance is 98% of the cross-race variance. This suggests a mixture of all three forces is present.

Our results should not be interpreted as saying that taste-based discrimination is absent or unimportant; our test can only identify at most two sources of discrimination even if all three are present. Instead, we show that screening discrimination seems to play an important role, especially when it comes to Hispanics. Thus, managers and policy makers should take screening discrimination into account when designing reforms (e.g. racial quotas, centralized hiring, and diverse hiring committees) that seek to identify the best applicants and reduce racial outcome gaps. Our results also suggest that the effectiveness and appropriateness of interventions also vary by minority group.

Finally, we present two extensions. Under a symmetry assumption, we show that we can use the three pairwise relations to identify the bias of each manager race. And we study the long-term implications of our results for racial composition of the firm's employees. In the Online Appendix we also discuss the sensitivity of our results to different regression specifications and other economic forces that may affect our results (in particular, customer preferences, peer effects, favoritism, referral networks, and worker preferences).

[4] For linguistic simplicity, we refer to white, black, and Hispanic all as "races."

1.1 Literature

There is widespread evidence for discrimination in recruiting. For example, correspondence studies find that applicants with white-sounding names receive higher callback rates than black-sounding names (Bertrand and Mullainathan, 2004; Kline, Rose, and Walters, 2022). Closer to our study, Giuliano, Leonard and Levine (2009, 2011) study hiring practices of managers of different races at a large chain of retail stores and find evidence of same-race bias between white, black and Hispanic employees (the latter in highly Hispanic locations). Our objective productivity measure allows us to go beyond these findings by partially identifying the cause of segregation.[5]

There is a large literature that uses lab and field experiments to identify labor market discrimination. For example, Hedegaard and Tyran (2018) find that Danish students prefer co-workers of the same ethnicity, even after productivity information is disclosed. Neumark (2018) has a wide-ranging survey on vignette studies and audit studies; these seek to separate taste-based discrimination from statistical discrimination by changing the characteristics of applicants or the information available to recruiters. Relative to these papers, we study hiring "in the wild" using administrative data.

A few recent papers have documented the extent and source of discrimination within the workplace. Glover, Pallais, and Pariente (2017) study a French grocery chain and show that minority cashiers are as productive as non-minority cashiers on average, but become less productive when working under managers who were deemed biased by an implicit association test. Their survey of workers indicates that biased managers neglect minority workers rather than treat them badly. Looking at a fruit picking farm, Bandiera, Barankay, and Rasul (2009) show that managers assist socially connected workers at the expense of unconnected workers when paid a fixed wage, but not when paid a performance bonus. And within a Kenyan flower packing firm, Hjort (2014) shows that workers try to lower the performance of team members from different tribes, even at a cost to themselves. These studies provide evidence of "complementary productivity," whereby workers are more productive when working with people of the same race. Relative to these papers, we will focus on the hiring decision and provide a new method to identify the source of discrimination. Our theory will draw a clear line between tased-based discrimination (managers prefer to hire less productive same-race applicants) and complementary productivity (managers make same-race workers more productive), but the underlying causes may be related. For example, animus may affect both the hiring decisions of managers and workers' productivity after being hired; our approach allows for the coexistence of both types of discrimination.

Another literature studies discrimination using matched employee-employer data.[6] Dustmann,

[5]Comparing our hiring results (Table 2) with Giuliano et al. (2009, Table 9), the black/white gap looks similar, but our white/Hispanic gap is larger. This may be because of different locations or demographics; e.g. in Giuliano et al. (2009, 2011), front-line workers are 70% female with an average age of 22, whereas our salespeople are 41% female with a median age of 32.

[6]There is also a large literature on statistical discrimination (whereby firms have more accurate signals of majority

Glitz, Schönberg, and Brücker (2016) show that firms in Germany with more minority employees hire more minority applicants and pay them slightly higher wages. They argue that these results come from referrals that provide information about match quality, which they interpret through a model of screening discrimination. Relatedly, Åslund, Hensvik, and Skans (2014) show that immigrant managers in Sweden are more than twice as likely to hire immigrants compared to native managers. Immigrants are also paid higher wages when managers are also immigrants; this is mainly explained by individual worker effects, suggesting that immigrant managers hire higher quality immigrant recruits, consistent with screening discrimination. Relative to these papers, our data concern relatively homogeneous jobs at a single US employer; we then show how the mean and variance of productivity can be used to distinguish different theories of discrimination.[7]

Methodologically, we are closest to Anwar and Fang (2006) who show how relative comparisons can be used to overcome the infra-marginality problem associated with Becker's (1957) outcome test. They study police officers' decisions to stop and search a driver and propose a simple test of relative taste-based discrimination. Specifically, police officers are prejudiced if the search rate exhibits rank reversal (e.g. white officers search more black drivers, while black officers search more white drivers). The test makes few distributional assumptions (e.g. about the underlying populations or police officers' private signals) but it is a rather weak test (e.g. if white police pull over many more people than black police then it is hard to detect bias). Our theory has a continuous outcome variable, which allows us to study both the mean and variance of productivity. This allows us to separate taste-based discrimination, screening discrimination and complementary production, whereas Anwar-Fang focus on taste-based discrimination. To study all these forces together we impose more parametric structure, in terms of normal distributions of priors and signals, and model heterogeneity via Taylor approximations. Given these assumptions, we say that managers are prejudiced if the mean output is submodular, which is less demanding than Anwar-Fang's rank reversal test.

The use of relative comparisons to evaluate discrimination has been used in other areas outside the employment domain. Ayres and Siegelman (1995) observe that black customers in car dealerships

applicants). Altonji and Pierret (2001) study the correlation between wages and intelligence scores as workers age, and find mixed evidence. Agan and Starr (2017) use audit methods to show that that ban-the-box laws that prohibit employers from asking about job applicants' criminal histories can lead employers to reduce the call-back rate for black applicants. Our model allows for workers to be drawn from different distributions, as in a traditional model of statistical discrimination, but our test concerns relative discrimination and thus only identifies screening discrimination.

[7]Screening discrimination has also been studied in the psychology and communication. For example, in an ethnographic study of 185 Chicago-area firms, Neckerman and Kirschenman (1991) find evidence that "interviewing well goes beyond interpersonal skills to common understandings of appropriate interaction and conversational style—in short, shared culture." More broadly, a literature on the contact hypothesis and social cognition finds that individuals think of other races categorically rather than as individuals, particularly when the focal person has limited meaningful experience interacting with members of the other group (Allport, Clark, and Pettigrew (1954), Amir (1969), Macrae and Bodenhausen (2000)). This is closely related to screening discrimination's key condition that hiring managers update their beliefs to a greater extent when candidates share their same race.

receive worse offers, no matter the race of the salesperson, which they interpret as evidence of statistical discrimination. List (2004) finds that baseball card traders believe minority customers have a wider variance in willingness-to-pay for baseball cards, leading them to offer minorities less favorable quotes. Fisman, Paravisini, and Vig (2017) looks at Indian loan decisions and shows that, consistent with screening discrimination, same-group borrowers receive more loans, more variable loan terms, and have lower default rates (where a "group" corresponds to religion or caste). In comparison to our formulation, Fisman et al. have a continuous decision variable (loan terms) and a binary outcome variable (default), whereas we study the mean and variance of a continuous outcome variable (productivity).

There are papers that seek to overcome the infra-marginality problem in different ways. Knowles, Persico, and Todd (2001) develop a model of the decision to carry contraband, in which marginal and average success rates of police searches will coincide. Anwar and Fang (2015) use the timing of release decisions of parole boards. Arnold, Dobbie, and Yang (2018) and Arnold, Dobbie, and Hull (2022) use variation between judges in bail decisions to infer the marginal treatment effect via two different methods; the latter paper seeks to identify taste-based and statistical discrimination. The bail data has the advantage that one observes each defendant (whereas we only see hired workers) and each judge makes a large number of decisions about randomly assigned cases. However, it uses a binary outcome variable (pretrial misconduct), whereas we study the mean and variance of a continuous outcome variable (productivity).

Looking across this expansive literature, most prior studies of labor market discrimination aim to provide evidence for individual theories. However, the effect of different policies depends on which theory is dominant. Our contribution is to provide an intuitive method to identify three different types of discrimination in one common framework.

2 Theory

We first describe a baseline model of recruiting based on Phelps (1972) and use it to examine the predictions that result from three theories of relative discrimination: taste-based discrimination, screening discrimination, and complementary production.[8] We then enrich the model to allow for manager and worker heterogeneity and propose our empirical tests.

Managers screen applicants for sales positions. Motivated by our data (see Figure B1 in Online Appendix B), we assume the log of workers' sales is normally distributed, $y \sim N(\mu, \sigma_0^2)$ and refer to y as the worker's *productivity*. A manager's *signal* is given by $\tilde{y} = y + \epsilon$, where the noise $\epsilon \sim N(0, \sigma_\epsilon^2)$

[8]With taste-based discrimination, a version of Proposition 1(a) was shown by Cornell and Welch (1996). With screening discrimination, Propositions 2(b),(d) are closely related to the analysis of Simon and Warner (1992) and Dustmann, Glitz, Schönberg, and Brücker (2016). We formalize these results and unify them in a common analytical framework.

is independent of y. Using Bayes' rule, the expected productivity of the applicant given the signal, the *estimate* $\hat{y}$, is a weighted sum of the prior and the signal, where the weights reflect the *signal variance* σ_ϵ^2 and the cross-sectional productivity variance σ_0^2,

$$\hat{y} = E[y|\tilde{y}] = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\epsilon^2}\tilde{y} + \frac{\sigma_\epsilon^2}{\sigma_0^2 + \sigma_\epsilon^2}\mu.$$

The estimate $\hat{y}$ is a sufficient statistic of the signal $\tilde{y}$ with a meaningful economic interpretation; so we work with $\hat{y}$ instead of $\tilde{y}$. As usual in Bayesian updating, it is important to distinguish between two variances. Ex-ante, the estimate $\hat{y}$ is distributed normally with mean μ and *estimator variance* $\eta^2 := \frac{\sigma_0^4}{\sigma_0^2+\sigma_\epsilon^2}$. Notably, the estimator variance η^2 *falls* in the signal variance σ_ϵ^2. For example, with a completely noisy signal, $\sigma_\epsilon^2 = \infty$, all workers have the same expected ability $\hat{y}$, so the estimator variance η^2 is zero. Ex-post, conditional on $\hat{y}$, the realized ability y is distributed normally with mean $\hat{y}$ and *residual variance* $\gamma^2 := \frac{\sigma_0^2\sigma_\epsilon^2}{\sigma_0^2+\sigma_\epsilon^2}$. For example, with a completely noisy signal, $\sigma_\epsilon^2 = \infty$, the residual variance equals the cross-sectional variance of productivity σ_0^2.

For ease of exposition, we assume that managers have log-utility so wish to hire a worker if expected log-sales $\hat{y}$ exceeds a cutoff y^*. The hiring probability is

$$p := \Pr[\hat{y} \geq y^*];$$

we also assume $y^* \geq \mu$, so at most half of all applicants are hired. Average productivity and productivity variance of accepted applicants are given by

$$\bar{y} := E[y|\hat{y} \geq y^*] \qquad \text{and} \qquad v := Var[y|\hat{y} \geq y^*].$$

Finally, we suppose the firm eventually learns the actual productivity y, and fires a worker if it is more than τ below the cutoff y^*. Turnover then equals

$$x := \Pr[y \leq y^* - \tau|\hat{y} \geq y^*].$$

We want to understand the effects of taste-based discrimination, screening discrimination, and complementary production on these four outcome variables $(p, \bar{y}, v, x)$. The proofs from this section are in Appendix A.

2.1 Taste-based discrimination

Under taste-based discrimination, managers are biased towards their own race. Specifically, they apply a lower hiring threshold to same-race hires than cross-race hires, $y_s^* < y_c^*$. They then fire workers whose productivity drops τ below the respective hiring threshold.

Proposition 1. *In the taste-based model:*

(a) The hiring probability is higher for same-race applicants than cross-race applicants, $p_s > p_c$*.*

(b) Average productivity is lower for same-race hires than cross-race hires, $\bar{y}_s < \bar{y}_c$*.*

(c) Productivity variance is higher for same-race hires than cross-race hires, $v_s > v_c$*.*

(d) Turnover is lower for same-race hires than cross-race hires, $x_s < x_c$*.*

The intuition is illustrated in Figure 1a (left). Part (a) says that managers are biased toward workers of their own race, so lower the required standard, and have a higher hiring probability, $p_s > p_c$. Part (b) shows that since managers apply a lower standard to same-race workers, their average productivity is lower. Part (c) says that the lower same-race standard $y_s^* < y_c^*$ raises the productivity variance of same-race recruits as additional low-productivity workers are hired. Specifically, by the law of total variance, we can decompose the productivity variance of recruits into the sum of estimator variance $Var(\hat{y}|\hat{y} \geq y_i^*)$ and residual variance $Var(y|\hat{y}) = \gamma^2$. The estimator variance is higher for same-race recruits, $Var(\hat{y}|\hat{y} \geq y_s^*) > Var(\hat{y}|\hat{y} \geq y_c^*)$, while the residual variance is equal for both races. Finally, part (d) says that turnover is lower for same-race hires. This result is somewhat counter-intuitive since same-race recruits have lower average productivity and higher variance. The result follows since managers apply the same bias when firing as when hiring. An agent is thus fired if the measurement error $\hat{y} - y$ exceeds the "safety margin" $\hat{y} - y^* + \tau$. Both same- and cross-race recruits have the same residual variance, but cross-race recruits have lower safety margins (in the likelihood ratio order). Indeed, in Figure 1a (left), one can see that cross-race recruits are closer to the hiring threshold than same-race recruits.

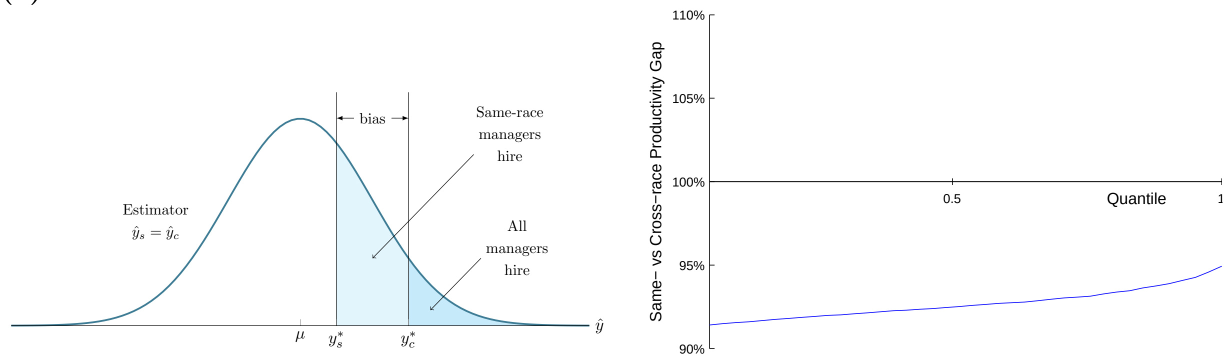
To illustrate the productivity results, Figure 1a (right) plots the difference between same-race and cross-race productivity by quantile. In the figure, quantile 0 corresponds to the difference between the worst same-race and cross-race recruits, while quantile 1 corresponds to the best recruits. One can see that the same-race recruits are less productive, as in part (b). This gap shrinks as the recruits get better since the best workers are hired no matter the race of the manager; this raises the variance for same-race hires, as in part (c).

2.2 Screening discrimination

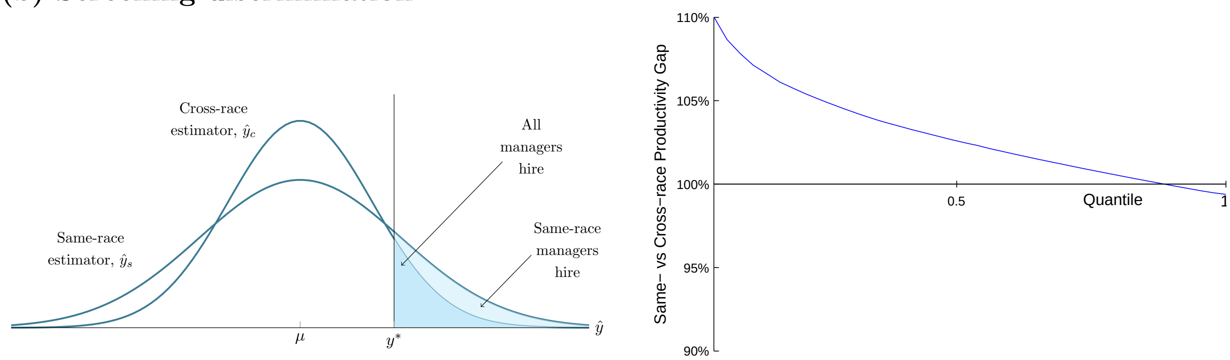
We now turn to screening discrimination, assuming signals are more accurate for same-race applicants than cross-race, so with signal variance $\sigma_s^2 < \sigma_c^2$. Managers are unbiased and hire applicants with expected productivity above threshold y^*, and fire workers whose realized productivity falls below $y^* - \tau$.[9]

[9] Given the different signal variances, one might argue that the ability to fire workers induces different cutoffs y^* for different races. This can go in either direction. If the managers take into account the option value of employment, the ability to fire low-productivity workers means managers would lower their standards for high-variance cross-race applicants. Conversely, if reversing one's hiring decision is viewed as a bad signal by the store manager, managers would raise their standards for cross-race applicants. For simplicity, we assume y^* is the same for each race.

(a) Taste-based discrimination



(b) Screening discrimination



(c) Complementary production

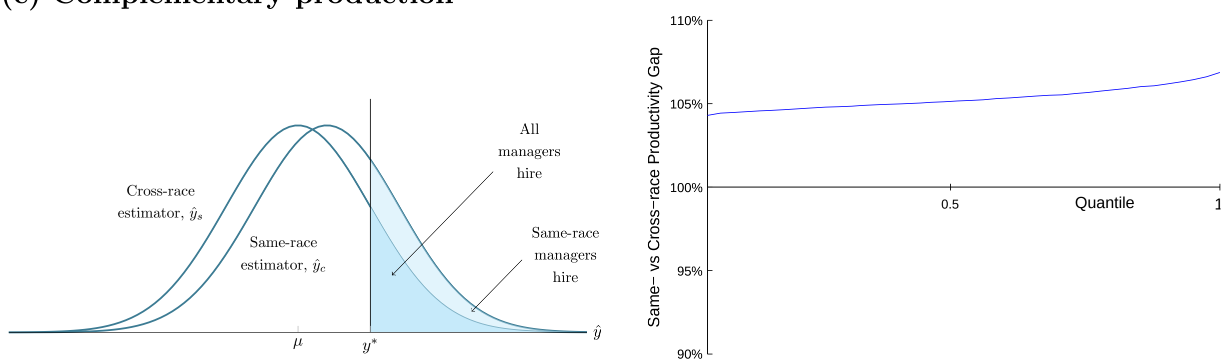


FIGURE 1: **DISTRIBUTION OF APPLICANTS' ESTIMATORS AND RECRUITS' PRODUCTIVITY**

Notes: The **left** figure illustrates the distribution of estimators $\hat{y}$ under same-race and cross-race managers. The **right** figure shows the difference between same-race and cross-race productivity by quantile (the y-axis shows the de-logged scale for ease of interpretation). These figures assume ability has mean $\mu = -0.5$ and variance $\sigma_0^2 = 0.4$, the signal variance is $\sigma_\epsilon^2 = 0.2$ and cutoff is $y^* = -0.3$; this generates a productivity distribution with mean 0.046, median 0.028 and variance 0.21, similar to our data (see Figure B1 in Online Appendix B). Under taste-based discrimination, the same- and cross-race cutoffs are $y_s^* = -0.3$ and $y_c^* = -0.2$. Under screening discrimination, the same- and cross-race signals have variance $\sigma_s^2 = 0.2$ and $\sigma_c^2 = 0.3$. Under complementary production, the same- and cross-race means are $\mu_s = -0.4$ and $\mu_c = -0.5$.

Proposition 2. *In the screening model:*

(a) The hiring probability is higher for same-race applicants than cross-race applicants, $p_s > p_c$*.*

(b) Average productivity is higher for same-race hires than cross-race hires, $\bar{y}_s > \bar{y}_c$*.*

(c) Productivity variance is lower for same-race hires than cross-race hires, $v_s < v_c$*.*

(d) Turnover is lower for same-race hires than cross-race hires, $x_s < x_c$*.*

The intuition is illustrated in Figure 1b (left). For part (a), managers have better signals about same-race applicants, so the estimator variance is higher and candidates have a higher probability of passing the hiring threshold. For part (b), note that since managers have better signals about same-race applicants, the estimator variance is higher. That is, there are many same-race candidates whose expected productivity $\hat{y}$ far exceeds the threshold y^*. In comparison, the estimated productivity of most cross-race hires is relatively close to y^*, and so same-race hires have higher expected productivity than cross-race hires. Part (c) states that the higher accuracy of the same-race signals means that the productivity variance is lower for same-race hires. Intuitively, if managers screen on a very noisy signal then the variance of realized productivity equals that of the prior; as the accuracy rises, more unproductive workers are excluded, and the variance of realized productivity falls. Finally, part (d) follows since same-race recruits have higher mean productivity and lower residual productivity variance. Both of these forces lower the chance of outliers, and thus turnover.

The predictions for hiring and turnover are the same as with taste-based discrimination (Proposition 1). Thus, hiring data alone is not enough to differentiate between the two models. To do this one needs productivity data, for which the models generate opposite predictions. To gain some further intuition for the productivity difference, Figure 1b (right) plots the difference between same-race and cross-race productivity by quantile. One can see that the gap is initially positive, shrinks as the workers get better, and eventually becomes negative. Intuitively, cross-race managers make some very bad hires since their signal is imprecise, leading to a large gap at the lowest quantiles, while everyone hires the best workers.[10]

2.3 Complementary production

We now turn to complementary production, assuming workers are more productive under same-race managers, $\mu_s > \mu_c$. For example, same-race managers may provide superior supervision, feedback, and training; this could arise because of cultural competency or bias of the manager.

Managers take into account the higher productivity of same-race workers, hiring any worker with expected productivity above threshold y^*; they fire any worker with realized productivity below threshold $y^* - \tau$.

[10] The cross-race manager also makes fewer hires overall, so the very best are a larger percentage of their hires. This explains why the productivity gap in Figure 1b (right) is negative at the highest quantiles.

Proposition 3. *In the complementary production model:*

(a) The hiring probability is higher for same-race applicants than cross-race applicants, $p_s > p_c$.

(b) Average productivity is higher for same-race hires than cross-race hires, $\bar{y}_s > \bar{y}_c$.

(c) Productivity variance is higher for same-race hires than cross-race hires, $v_s > v_c$.

(d) Turnover is lower for same-race hires than cross-race hires, $x_s < x_c$.

The intuition is illustrated in Figure 1c (left). For part (a), same-race workers are more productive, so managers hire more of them. For parts (b)-(d), the higher productivity increases the proportion of infra-marginal same-race workers with productivity far above the threshold. Thus, the productivity of same-race workers has higher mean and variance, and turnover is lower. To illustrate the productivity differences, Figure 1c (right) plots the difference between same-race and cross-race productivity by quantile. One can see that the gap is positive and grows as workers get better. Intuitively, complementarity is at work at the top of the distribution, whereas it is offset by selection at the bottom.

2.4 Discrimination tests

In Sections 2.1-2.3, we consider one theory at a time. To bring our model to the data we nest the three theories in one common model that additionally allows for worker and manager heterogeneity by race. This analysis directly leads to our supermodularity tests.

General model. Consider a population containing mass n_Θ race-Θ workers where $\Theta \in \{W, B, H, ...\}$, that are screened by race-θ mangers where $\theta \in \{w, b, h, ...\}$. We write $\Theta = \theta$ when, say, $\Theta = W$ and $\theta = w$. The model is indexed by parameters $(\chi_\Theta, \xi_{\theta'}; d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'})$ that come from two components. First, we model worker heterogeneity by parameters χ_Θ and manager heterogeneity by parameters ξ_θ. For workers, χ_Θ could reflect differences in the population mean μ or variance σ_0^2. For managers, ξ_θ could reflect differences in hiring standards y^* or signal quality σ_ϵ^2. Second, we incorporate three forms of discrimination. Taste-based discrimination is indexed by parameters $d_{\Theta\theta'} \geq 0$, which additively enter the hiring threshold $y^*_{\Theta\theta'} = y^* + d_{\Theta\theta'}$. Screening discrimination is indexed by parameters $z_{\Theta\theta'} \geq 0$, which multiplicatively enter the signal variance $\sigma^2_{\Theta\theta'} = (\sigma_\epsilon(1 + z_{\Theta\theta'}))^2$. Complementary production is indexed by parameters $k_{\Theta\theta'} \geq 0$, which additively enters the population productivity average $\mu_{\Theta\theta'} = \mu - k_{\Theta\theta'}$.[11] We normalize $(d_{\Theta\theta}, z_{\Theta\theta}, k_{\Theta\theta}) = (0, 0, 0)$ for worker-manager pairs of the same race. When $(d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'}) = (0, 0, 0)$ then race-θ' managers do not discriminate against race-Θ workers.

We wish to understand how heterogeneity and discrimination $(\chi_\Theta, \xi_{\theta'}; d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'})$ affect our output variables $(p, \bar{y}, v, x)$. We trace out the impact of the parameters via Taylor expansions.

[11] Note that k enters as a negative term in the productivity of cross-race workers, so k more literally measures "mismatch" rather than "complementarity". We do this so we can allow for heterogeneity across race-pairs and because taste-based discrimination d and screening discrimination z are also formulated for of cross-race hires.

For example, the productivity mean of race-Θ workers equals $\bar{y}_{\Theta\theta} = \bar{y}(\chi_\Theta, \xi_\theta, 0, 0, 0)$ when hired by same-race managers $\theta = \Theta$, and $\bar{y}_{\Theta\theta'} = \bar{y}(\chi_\Theta, \xi_{\theta'}, d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'})$ when hired by cross-race managers $\theta' \neq \Theta$. Taking a Taylor approximation around $(\chi_\Theta, \xi_{\theta'}; d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'}) = \mathbf{0}$,

$$\bar{y}_{\Theta\theta'} = \bar{y}(\mathbf{0}) + \chi_\Theta \frac{\partial \bar{y}}{\partial \chi} + \xi_{\theta'} \frac{\partial \bar{y}}{\partial \xi} + d_{\Theta\theta'} \frac{\partial \bar{y}}{\partial d} + z_{\Theta\theta'} \frac{\partial \bar{y}}{\partial z} + k_{\Theta\theta'} \frac{\partial \bar{y}}{\partial k} \tag{1}$$

We can similarly take Taylor approximations for the other outcome variables. Echoing Table 1, Propositions 1-3 then tell us about their partial derivatives,

$$\begin{array}{rllll} \text{Taste-based discrimination:} & \dfrac{\partial p}{\partial d} < 0, & \dfrac{\partial \bar{y}}{\partial d} > 0, & \dfrac{\partial v}{\partial d} < 0, & \dfrac{\partial x}{\partial d} > 0 \\ \text{Screening discrimination:} & \dfrac{\partial p}{\partial z} < 0, & \dfrac{\partial \bar{y}}{\partial z} < 0, & \dfrac{\partial v}{\partial z} > 0, & \dfrac{\partial x}{\partial z} > 0 \\ \text{Complementary productivity:} & \dfrac{\partial p}{\partial k} < 0, & \dfrac{\partial \bar{y}}{\partial k} < 0, & \dfrac{\partial v}{\partial k} < 0, & \dfrac{\partial x}{\partial k} > 0 \end{array} \tag{2}$$

The underlying assumption behind the use of Taylor expansions is that the heterogeneity and bias are small. This assumption seems reasonable given that mean productivity of different race pairs varies by at most 4% (see Table B1 in Online Appendix B); these numbers capture both worker and manager heterogeneity (χ, ξ) and the discrimination parameters (d, z, k).[12]

Tests of discrimination. Using our general model, we seek to measure the form and extent of discrimination. One might first think about comparing one type of worker under different types of managers (e.g. asking whether Hispanic workers are more productive under white or Hispanic managers). However, this "within worker test" is undermined by the managerial heterogeneity ξ_θ. One can see this by using our Taylor approximation (1) to compare the productivity of worker Θ under same-race manager θ and cross-race manager θ',

$$\begin{aligned} \bar{y}_{\Theta\theta} - \bar{y}_{\Theta\theta'} &= \bar{y}(\chi_\Theta, \xi_\theta, 0, 0, 0) - \bar{y}(\chi_\Theta, \xi_{\theta'}, d_{\Theta\theta'}, z_{\Theta\theta'}, k_{\Theta\theta'}) \\ &= \underbrace{(\xi_\theta - \xi_{\theta'}) \frac{\partial \bar{y}}{\partial \xi}}_{\text{Manager Heterogeneity}} \underbrace{- d_{\Theta\theta'} \frac{\partial \bar{y}}{\partial d} - z_{\Theta\theta'} \frac{\partial \bar{y}}{\partial z} - k_{\Theta\theta'} \frac{\partial \bar{y}}{\partial k}}_{\text{Discrimination Effects}}. \end{aligned}$$

For example, Hispanic workers may be more productive under Hispanic managers than under white managers because Hispanic managers have a large positive fixed effect, ξ_θ.[13]

To overcome the problem of worker and manager heterogeneity, we take a diff-in-diff approach

[12] A caveat: Taylor approximations are valid for "small" perturbations but it is not generally possible to quantify the meaning of "small" without using the structure of our problem. Given the five types of heterogeneity in our model, studying the full model would quickly become intractable.

[13] One can similarly compare the productivity of different races of workers under the same manager. But the sign of this "within manager effect" may simply reflect differences in worker populations, $(\chi_\Theta - \chi_{\Theta'})\frac{\partial \bar{y}}{\partial \chi}$.

and propose a *supermodularity test* of mean productivity. Define

$$\Delta^y_{\Theta\Theta'} := \underbrace{(\bar{y}_{\Theta\theta} - \bar{y}_{\Theta\theta'})}_{\text{Productivity gap for } \Theta} - \underbrace{(\bar{y}_{\Theta'\theta} - \bar{y}_{\Theta'\theta'})}_{\text{Productivity gap for } \Theta'} \tag{3}$$

where $\Theta = \theta$ and $\Theta' = \theta'$.[14] For example, this compares the productivity boost Hispanic workers get from having a Hispanic manager to the productivity boost that white workers get from having a Hispanic manager. Using the Taylor expansion (1),

$$\Delta^y_{\Theta\Theta'} = -(d_{\Theta\theta'} + d_{\Theta'\theta})\frac{\partial \bar{y}}{\partial d} - (z_{\Theta\theta'} + z_{\Theta'\theta})\frac{\partial \bar{y}}{\partial z} - (k_{\Theta\theta'} + k_{\Theta'\theta})\frac{\partial \bar{y}}{\partial k}. \tag{4}$$

Recall from (2) that $\partial \bar{y}/\partial d$ is positive, whereas $\partial \bar{y}/\partial z$ and $\partial \bar{y}/\partial k$ are negative. Therefore, if mean output is submodular, $\Delta^y_{\Theta\Theta'} < 0$, then $d_{\Theta\theta'} + d_{\Theta'\theta} < 0$ which means at least one race of manager uses taste-based discrimination.

We next consider productivity variance. Analogous to the classic "ratio of variances" test, we define the log-supermodularity test of productivity variance,

$$\Delta^v_{\Theta\Theta'} := \frac{v_{\Theta\theta}/v_{\Theta\theta'}}{v_{\Theta'\theta}/v_{\Theta'\theta'}} \tag{5}$$

Taking logs and applying the Taylor approximation

$$\begin{aligned}\log \Delta^v_{\Theta\Theta'} &= (\log v_{\Theta\theta} - \log v_{\Theta\theta'}) - (\log v_{\Theta'\theta} - \log v_{\Theta'\theta'}) \\ &= \frac{1}{v}\left(-(d_{\Theta\theta'} + d_{\Theta'\theta})\frac{\partial v}{\partial d} - (z_{\Theta\theta'} + z_{\Theta'\theta})\frac{\partial v}{\partial z} - (k_{\Theta\theta'} + k_{\Theta'\theta})\frac{\partial v}{\partial k}\right)\end{aligned}$$

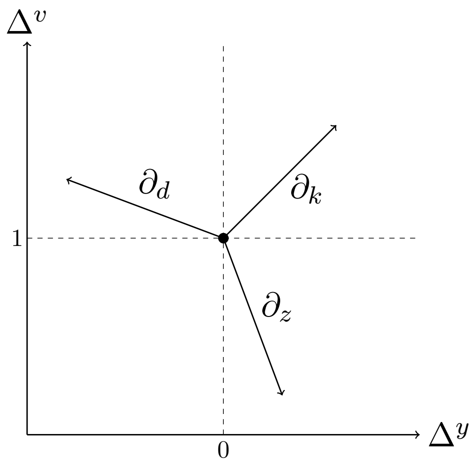
By (2), $\partial v/\partial z$ is positive, whereas $\partial v/\partial d$ and $\partial v/\partial k$ are negative. Therefore, if the variance is log-submodular, in that $\log \Delta^v_{\Theta\Theta'} < 0$ or $\Delta^v_{\Theta\Theta'} < 1$, then $z_{\Theta\theta'} + z_{\Theta'\theta} > 0$ which means at least one race of manager uses screening discrimination.

Using these insights, we can use productivity data to identify all three discrimination theories:

Proposition 4 (**Tests of discrimination**). *Consider the general model with worker and manager heterogeneity, and assume the Taylor approximations are valid.*
(a) If $\Delta^y_{\Theta\Theta'} < 0$*, then there exists taste-based discrimination:* $d_{\Theta'\theta} + d_{\Theta\theta'} > 0$.
(b) If $\Delta^v_{\Theta\Theta'} < 1$*, then there exists screening discrimination:* $z_{\Theta'\theta} + z_{\Theta\theta'} > 0$.
(c) If $\Delta^y_{\Theta\Theta'} > 0$ *and* $\Delta^v_{\Theta\Theta'} > 1$*, then there exists complementary production:* $k_{\Theta'\theta} + k_{\Theta\theta'} > 0$.

Part (c) says that there is no combination of taste-based and screening discrimination that could explain $\Delta^y_{\Theta\Theta'} > 0$ and $\Delta^v_{\Theta\Theta'} > 1$. Intuitively, taste-based discrimination most directly affects expected productivity, while screening discrimination most directly affects productivity variance.

[14] We use the double-capitalized notation $\Theta\Theta'$ to indicate a general pair of races, rather than a worker-manager race pair $\Theta\theta'$.



	$\Delta^y < 0$	$\Delta^y > 0$
$\Delta^v > 1$	Taste-based disc.	Complementary prod.
$\Delta^v < 1$	Screening disc. & Taste-based disc.	Screening disc.

FIGURE 2: **IDENTIFYING DISCRIMINATION FROM PRODUCTIVITY DATA**

Notes: The figure on the **left** illustrates how taste-based discrimination, screening discrimination and complementary production affect the supermodularity of mean productivity Δ^y and productivity variance Δ^v. By (6), ∂_z is steeper than ∂_p. The matrix on the **right** shows the table we will use for our empirical tests.

Formally, we show in Appendix A.2

$$\frac{-\partial v/\partial d}{\partial \bar{y}/\partial d} < \frac{\partial v/\partial z}{-\partial \bar{y}/\partial z}. \tag{6}$$

Figure 2 (left) illustrates the impact of the three types of discrimination (d, z, k) on (Δ^y, Δ^v). In particular, it shows that an increase in any one variable (d, z, k) cannot be replicated by an increase in the other two variables. The matrix in Figure 2 (right) then summarizes our tests.

One may wonder whether Proposition 4 captures all empirical implications of our model: Are there additional combinations of (Δ^y, Δ^v) that allow us to conclude that one of our discrimination forces must be at play? For screening discrimination this is not the case. We show in Appendix A.2 that the slope of any of the three vectors in Figure 2 is linear in the estimator variance η, which is unobservable, and hence potentially arbitrarily small. For small η, the cone spanned by ∂_d and ∂_k covers the entire upper half-plane, so Proposition 4(b) is tight. However, the complementarity test, Proposition 4(c), can be tightened: The slope of ∂_d is bounded above, so the triangle area above ∂_d in the upper-left quadrant can only be reached with ∂_k, and is hence proof of complementary production. We do not pursue this angle further in this paper since such considerations would not substantially change our empirical conclusions.[15]

Tests of the model. Figure 2 implies that linear combinations of ∂_d, ∂_z, and ∂_k with positive

[15] The above discussion concern restrictions on individual vectors $(\partial_d, \partial_z, \partial_k)$. We also know that there exists joint restrictions on these vectors. In particular, Figure 2 (left) shows that there always exists a half-space to the upper right of the origin that is proof of complementary production, although we do not know which half-space this is. If 95% of our bootstraps lie in all half-planes to the upper left of the origin, we could conclude there is complementary production; this is a more powerful test than the one proposed in Proposition 4(c), where we just look at the upper right quadrant. For our data, this does not affect our conclusions.

coefficients span the entire plane; in other words productivity data cannot reject the joint model. However, all three discrimination theories have the same predictions for hiring and turnover. This gives us a way of rejecting our model, even in the presence of worker and manager heterogeneity.

Hiring probabilities $p_{\Theta\theta'}$ are not empirically observable. We address this problem by measuring the share of race-Θ workers among the $N_{\theta'} = \sum_{\theta''} n_{\Theta''} p_{\Theta''\theta'}$ workers hired by race-θ' managers, $r_{\Theta\theta'} = \frac{n_\Theta p_{\Theta\theta'}}{N_{\theta'}}$. As with productivity variance we define the log-supermodularity test of hiring,

$$\Delta^r_{\Theta\Theta'} := \frac{r_{\Theta\theta}/r_{\Theta\theta'}}{r_{\Theta'\theta}/r_{\Theta'\theta'}} \tag{7}$$

Our theory predicts that $\Delta^r_{\Theta\Theta'} > 1$. To see this, a Taylor approximation and the inequalities in (2) yield,

$$\begin{aligned}\log \Delta^r_{\Theta\Theta'} &= (\log r_{\Theta\theta} - \log r_{\Theta\theta'}) - (\log r_{\Theta'\theta} - \log r_{\Theta'\theta'}) \\ &= \frac{1}{p}\left(-(d_{\Theta\theta'} + d_{\Theta'\theta})\frac{\partial p}{\partial d} - (z_{\Theta\theta} + z_{\Theta'\theta})\frac{\partial p}{\partial z} - (k_{\Theta\theta} + k_{\Theta'\theta})\frac{\partial p}{\partial k}\right) > 0.\end{aligned}$$

Finally, we define the supermodularity test of turnover

$$\Delta^x_{\Theta\Theta'} := (x_{\Theta\theta} - x_{\Theta\theta'}) - (x_{\Theta'\theta} - x_{\Theta'\theta'}). \tag{8}$$

Our theory predicts that $\Delta^x_{\Theta\Theta'} < 0$. To see this, a Taylor approximation and the inequalities in (2) yield,

$$\Delta^x_{\Theta\Theta'} = -(d_{\Theta\theta'} + d_{\Theta'\theta})\frac{\partial x}{\partial d} - (z_{\Theta\theta} + z_{\Theta'\theta})\frac{\partial x}{\partial z} - (k_{\Theta\theta} + k_{\Theta'\theta})\frac{\partial x}{\partial k} < 0.$$

We have thus shown

Proposition 5 (Tests of the model). *Consider the general model with worker and manager heterogeneity, and assume the Taylor approximations are valid.*
(a) Hiring is log-supermodular, $\Delta^r_{\Theta\Theta'} > 1$.
(b) Turnover is submodular, $\Delta^x_{\Theta\Theta'} < 0$.

Part (a) states that the ratio of workers hired by manager θ, $r_{\Theta\theta}/r_{\Theta'\theta}$, exceeds the ratio hired by manager θ', $r_{\Theta\theta'}/r_{\Theta'\theta'}$. One might wonder whether, more strongly, each manager will hire a greater proportion of same-race workers, $r_{\Theta\theta} > r_{\Theta\theta'}$ and $r_{\Theta'\theta'} > r_{\Theta'\theta}$. This need not be the case: If, say, white managers strongly discriminate against Hispanic workers then they may end up hiring more black workers than a black manager. But the ratio of white/black workers will always be higher for the white manager.

2.5 Discussion

We first discuss our tests and then turn to the robustness of the theoretical results.

Power of the test. Our test will not always detect discrimination, even as the sample size grows to infinity. If all three forms of discrimination are present, then we can detect at most two (see Figure 2). Indeed, if the different forms of discrimination offset one another, our test may be unable to identify any of them. In addition, our test can detect how managers of different races treat employees ("relative discrimination") but cannot tells us if all managers are biased against a particular race of worker ("absolute discrimination").

Random assignment. Our key identifying assumption is that workers are randomly assigned to managers. Formally, this means that the number of applicants n_Θ and the distribution of worker skills $(\mu_\Theta, \sigma^2_\Theta)$ are independent of manager race, θ. Several forces may conflict with this assumption:

- Regional variation. Our data contains stores across the US, and applicants at a given store may vary in their racial mix, n_Θ, and their skill distribution $(\mu_\Theta, \sigma^2_\Theta)$. Since manager racial composition also varies across stores, we control for store fixed effects.

- Job differentiation. Jobs in some departments may be harder and higher paid than others, attracting different numbers of applicants n_Θ and skill distributions $(\mu_\Theta, \sigma^2_\Theta)$. Departments may differ in their managerial race composition, so we control for department fixed effects.

- Intertemporal changes. As the retail sector changes over time (because of seasonality or sectoral decline), the composition of applicants may change in their racial mix, n_Θ, and their skill distribution $(\mu_\Theta, \sigma^2_\Theta)$. Managerial race composition may also change over time, so we control for month fixed effects.

Ultimately, the hiring analysis (Proposition 5(a)) assumes each manager faces an identical distribution of workers in terms of the number of applicants n_Θ and distribution of skills $(\mu_\Theta, \sigma^2_\Theta)$ after controlling for fixed effects. The test of discrimination (Proposition 4) and turnover (Proposition 5(b)) just require that each manager faces an identical distribution of skills $(\mu_\Theta, \sigma^2_\Theta)$ after controlling for fixed effects; differences in the number of workers n_Θ do not change the sub/supermodularity of productivity or turnover.

Comparison with outcome test. The classic outcome test (Becker, 1957) examines whether a firm is biased by comparing the performance of marginal applicants. We do not observe the managers' private signals and thus do not observe the marginal worker. A naive comparison of mean productivity $E_\theta[\bar{y}_{\Theta\theta}]$ across workers Θ would then suffer from both false positives and false negatives. Specifically, if worker Θ were more productive than Θ' because of higher fixed effects χ_Θ (e.g. a higher prior mean), we would see $E_\theta[\bar{y}_{\Theta\theta}] > E_\theta[\bar{y}_{\Theta'\theta}]$ even if there were no discrimination. Conversely, if two equally sized populations exhibit symmetric discrimination (e.g. $d_{\Theta\theta'} = d_{\Theta'\theta} > 0$) then we could see $E_\theta[\bar{y}_{\Theta\theta}] = E_\theta[\bar{y}_{\Theta'\theta}]$ even in the presence of discrimination. Rather than try to identify the marginal recruit, we follow a second approach pioneered by Anwar and Fang (2006) of looking for relative bias between managers of different races.

Hiring decision. The model assumes that managers have log utility, which means that they hire a worker if expected log-sales $\hat{y}$ exceeds a cutoff, y^*. Instead, suppose managers are risk neutral and wish to maximize expected sales; they are then risk-loving with respect to log-sales. Under taste-discrimination and complementary production the analysis is unchanged. Under screening discrimination, things become a little more complicated since managers now apply a *lower* hiring threshold to cross-race hires than same-race hires because of the higher residual variance of cross-race applicants. This new force works against the bias for hiring same-race candidates, Proposition 2(a), but the result still holds if the hiring threshold is sufficiently high. On the other hand, the lower threshold means managers hire more low-productivity cross-race workers, and thus reinforces Propositions 2(b,c,d). See Online Appendix C.1 for details.

The model also assumes that managers use a deterministic cutoff y^*. This cutoff may be random if managers have varying standards or if workers compete for jobs in tournaments.[16] Allowing y^* to have an arbitrary distribution does not affect the results for hiring and firing in parts (a) and (d) of Propositions 1–3. The analysis of productivity mean and variance have additional terms which are hard to sign, but by continuity the comparative static in parts (b) and (c) of Propositions 1–3 are still valid if the noise is small (e.g. there are a large number of applicants and open positions).

Firing decision. The model assumes an agent is fired if their realized productivity is below $y^* - \tau$. Under all three theories, turnover is submodular, so turnover cannot be used to identify the source of discrimination. One might have concerns with this modeling assumption. First, under taste-discrimination, it means that same-race workers have a lower firing bar than cross-race workers. Such firing discrimination may not occur if firing decisions are based on observable sales or if there is managerial turnover. A "neutral" firing bar would result in higher rates of turnover for same-race hires than cross-race hires, $x_s > x_c$, and would provide another way to identify the source of discrimination. A second concern is that some turnover is due to workers voluntarily leaving rather than being fired. In our data, "leavers" have lower productivity than "stayers", but retention could depend on workers' preferences in addition to their productivity. Given these questions about about the "correct" model, we think of our current model as the most conservative one since it says that one cannot use turnover data to identify the source of discrimination. Indeed, Propositions 1(d)-3(d) provide a cautionary tale for those tempted to use turnover as a proxy for productivity.

[16] To see how tournaments are equivalent to a random cutoff, note that worker i wins if discrimination-adjusted estimate of his productivity, $\hat{y}_i - d_i$, exceeds all that of all the other workers $\{\hat{y_1} - d_1, \hat{y_2} - d_2, \ldots\} := y^*$. If there are a Poisson number of applicants (motivated by the law of small numbers), then the hiring threshold y^* is independent of i's race.

3 Empirical Setting and Approach

Our data come from the U.S. operations of a large national retailer from February 2009 to March 2015. Each establishment is led by a store manager and a team of department managers (henceforth "managers") who hire for their respective departments, among other duties.

Hiring and employment. When a department has a vacancy, the manager requests a shortlist of qualified applicants from the regional HR representative, who assembles a list from online advertisements specific for that posting or from qualified "evergreen" applicants who can apply for entry-level positions at any time. Applicants to sales roles take a screening test online or at a store location, which generally requires less than an hour to complete. The test asks candidates about their background, technical qualifications, experience, how they would respond under hypothetical scenarios with customers or colleagues, and a personality test. These tests are scored by an algorithm that provides a three-tier recommendation for whether to proceed with an interview. The department managers observe the algorithmic recommendations and select whom to interview; they are not obligated to follow the recommendation.

Job interviews can vary substantially, but managers receive a guide for conducting behavioral interviews that further ask applicants how they would respond to hypothetical scenarios. Consistent with this training, most employees report that managers use behavior-based questions. For example, the applicant may be asked to describe a life or work experience in which they overcame an obstacle, helped a stranger, or witnessed dishonesty. After being hired, a worker typically undergoes about one week of formal online training and a week of job shadowing before moving to regular status. Salespeople are highly incentivized to maximize their productivity because commissions and other forms of incentive pay account for about 40% of their income, with the bulk of this pay tied directly or indirectly to their individually-credited sales performance.

Department managers have many duties outside hiring. They supervise salespeople, schedule shifts, give feedback to employees, handle customer complaints, and so on. They also play a significant role running the whole store, e.g. opening and closing the store, helping the manager prepare for sales events. These managers are also highly incentivized to maximize the productivity of their team since they are assessed and rewarded on the sales performance of their departments, which depends on the cumulative sales performance of their team. In particular, the company uses annual bonuses (averaging 5-10% of salary, but potentially much higher), retention offers, and promotion to encourage managers to hire productive salespeople and ultimately build a profitable department.

Descriptive statistics. The data we use for our main analyses include longitudinal administrative records on 63,842 commissioned salespeople and their managers, including productivity, demographic identity, department, and store location. Of these, 56,071 have an observable hiring manager because they are either directly observed at their hire date (47,025) or have a known hiring manager due to available position tenure data (9,046). To focus on racial combinations for whom we have the greatest statistical power, we restrict the sample to white, black, and Hispanic (WBH) workers hired by WBH managers. The resulting sample consists of 48,755 salespeople hired by 7,892 managers at 997 store locations over 74 months, or 335,867 worker-months. Salespeople are 64% white, 19% black, and 17% Hispanic, whereas managers are 77% white, 11% black, and 12% Hispanic. The age quartiles of salespeople are (21, 25, 42), the mean age is 32, and 41% are female. For department categorization, we use the 44 departments with at least 1,000 person-months in the full sample; the remaining 3.8% of salespeople are in an "other" category.

Our first variable of interest is the *new hire race*. As one would expect, there is a fair degree of regional segregation: white managers' WBH hires are 71.5% white, black managers WBH hires are 43.9% black, and Hispanic managers' WBH hires are 48.7% Hispanic. Fortunately, there is still a lot of variation of race within stores. Of the 997 stores, 423 have white and black managers, 412 have white and Hispanic managers, 217 have black and Hispanic managers, and 214 stores have all three. Nearly all (98.9%) new commissioned salespeople were hired into locations with variation in the race of new commissioned salespeople.

Our second variable of interest is *sales productivity*. We begin with a worker's sales per hour as a percent of their sales targets, which serves as the basis for their commission pay. Sales targets depend on several factors including shift, location, department, period, and tenure (but not race). Targets are set centrally by corporate headquarters and not by the managers. As shown in Figure B1 in Online Appendix B, the sales divided by the target distribution is roughly log-normal, so we use the performance measure $SPH_{i,t} = \ln(\text{sales per hour/target})$, and winsorize this variable at 1% within race pairs. To be consistent with the theory, we restrict the data to workers who work under their hiring manager. The resulting sales productivity measure has a mean of 0.007, reflecting that the average worker hits their target almost exactly. It has a standard deviation 0.518, reflecting substantial variation in sales performance.[17] This performance measure features small differences in the mean productivity of salespeople/managers of different races (see Table B1 in Online Appendix B). These could come from heterogeneity in the underlying population or from compositional effects in the model (e.g. under screening discrimination or complementary production, the minority race has lower mean productivity simply because they are in the minority). Our supermodularity tests are robust to such heterogeneity.

The third variable of interest is *turnover*. We first remove 2,703 salespeople who were hired before the start of our data and 6,713 salespeople who were hired within six months of the end of the data, leaving us with 39,339 people. After six months, 32.6% leave the firm directly from sales,

[17] Most of this variance comes from "across workers" rather than "within worker". The standard deviation of the worker-average residuals, $\bar{e}_i$, from the baseline productivity regression (11) equals 0.439.

3.5% switch to a non-sales job, and 59.7% stay within sales. Our theory posits that a worker is fired if the output drops below a cutoff; consistent with this, the average SPH is -0.0129 for leavers, 0.0036 for switchers, and 0.0256 for stayers.[18] Given the switchers perform worse than the leavers, we pool the two groups and say someone has turned over if they leave sales. Turnover is then 26.4% after 3 months, 36.1% after 6 months, 44.5% after 9 months and 51.7% after 12 months.

Data features and limitations. The data have a number of features that make it well suited to study discrimination in hiring. First, the nature of the job means that productivity largely reflects individual performance. In comparison, in many other industries (e.g. manufacturing) and even other sales jobs (e.g. complex business-to-business sales), performance reflects the work of an entire team. Second, as discussed above, salespeople and managers are highly incentivized to maximize their productivity. Third, managers have a large degree of autonomy in assessing applicants, allowing us to identify the impact of managerial race. Fourth, the firm is demographically and educationally similar to the greater retail trade industry, which employs over 16 million workers (more than 10% of the US labor force). These jobs are also similar in skill and labor force attachment to those held by many working class Americans where racial disparities seem to be largest (Lang and Lehmann, 2012). Finally, this firm employs about 1.2 million workers over the seven year period, affording a large amount of data on demographics, hiring, and turnover, as well as perhaps the largest such data on commissioned workers. Similarly sized data, like the American Community Survey's sample of employed workers (which covers about 1% of all U.S. workers each year), are missing key variables like job performance and manager race.

The main data limitation is that we cannot test the random assignment assumption. As discussed in Section 2.5, the model assumes that the number of applicants n_Θ and their productivity $(\mu_\Theta, \sigma^2_\Theta)$ is independent of the manager's race θ. Given that regional HR compiles the list of applicants, we believe that this is a reasonable assumption, but it is ultimately untestable within our data set. In particular, we do not observe the number or race of applicants, which would help us test the random assignment of worker race and provide data about the probability of acceptance.[19] And we do not observe the company's three-tier recommendation, which would help us test for the random assignment of worker productivity. Another limitation of the data is that we do not observe the shift assignments, but the target does control for shift since "sales per hour/target" is the basis for salespeople's commission.

Our baseline results examine hiring and productivity for newly hired salespeople. In Online Appendix B.1, we assess the hiring and turnover hypotheses with the full set of employment records, rather than just salespeople, greatly enhancing our sample size. Such workers include cashiers,

[18]The productivity measures come from the worker-average residuals, $\bar{e}_i$, from the baseline productivity regression (11).

[19]One natural concern is the prevalence of referrals. An employee within HR told us that it is rare that department managers know applicants personally, but we cannot rule it out. We discuss referrals further in Online Appendix B.2.

backroom operators, and service technicians. In all, there are 812,399 WBH workers at 4,646 store locations, which may be the largest such employment data. Of these workers, 57% are white, 25% are black, and 18% are Hispanic. The turnover is 26.5% at 3 months, 49.2% at 6 months, 62.2% at 9 months, and 70.6% at 12 months.

3.1 Empirical implementation

Our analysis is based on three regressions.

Hiring. Our model predicts that managers hire more same-race workers no matter the source of discrimination (see Proposition 5(a)). To test this we run the following regression for each worker race Θ,[20]

$$100 \times \text{Worker}_\Theta = \beta^r_{\Theta\theta}\text{Manager}_\theta + \text{controls}_{i,t} + \epsilon^r_i \tag{9}$$

On the LHS, we have an indicator that tells us whether the worker's race is Θ multiplied by 100. On the RHS, we have indicators that tell us whether that worker's manager is θ. The *controls* consist of 997 store fixed effects, 45 department fixed effects, and 74 month fixed effects. Thus, $\beta^r_{\Theta\theta}$ equals the adjusted share of a θ managers' employees who are of race Θ. Based on (7), define

$$\hat{\Delta}^r_{\Theta\Theta'} = \frac{\hat{\beta}^r_{\Theta\theta}/\hat{\beta}^r_{\Theta\theta'}}{\hat{\beta}^r_{\Theta'\theta}/\hat{\beta}^r_{\Theta'\theta'}} \tag{10}$$

where $\Theta = \theta$ and $\Theta' = \theta'$. Under the assumption that workers are randomly assigned to managers after controlling for fixed effects, Proposition 5(b) predicts that $\hat{\Delta}^r_{\Theta\Theta'} > 1$. That is, the ratio of white/Hispanic workers hired by white managers exceeds the ratio hired by Hispanic managers.

Productivity. We use the mean and variance of productivity to tease apart the three types of discrimination (see Proposition 4). To perform our productivity test, we consider the random effects model

$$SPH_{i,t} = \beta^y_{\Theta\theta} + \text{controls}_{i,t} + \ln(\text{tenure})_{i,t} + u_i + \epsilon^y_{i,t} \tag{11}$$

for $i = 1 \ldots N_{\Theta\theta}$ and $t = 1 \ldots T_i$, where $N_{\Theta\theta}$ are the number of Θ workers under θ managers and where T_i are the number of observations of worker i. We allow for the same controls as in (9) and additionally control for tenure. Agent i's productivity $u_i \sim N(0, \sigma^2_{\Theta\theta})$ and the monthly noise $\epsilon^y_{i,t} \sim N(0, \sigma^2_{\epsilon,\Theta\theta})$ are IID.[21] Thus, $\beta^y_{\Theta\theta}$ is the mean productivity of worker Θ under manager θ, and $\sigma^2_{\Theta\theta}$ is the variance of worker Θ under manager θ.

[20] Our findings are the qualitatively the same for logit or probit regressions. The interpretation is easier with a linear regression.

[21] In the model, the prior distribution of productivity is normal, however the posterior is truncated based on a signal and thus technically not normal. Empirically productivity looks normal (see Figure B1 in Online Appendix B), which is consistent with the model if managers' signals are fairly weak.

We estimate the mean productivity of $\Theta\theta$ pairs, $\beta^y_{\Theta\theta}$, via weighted least squares, where each worker is weighted by the inverse of T_i.[22] We inverse-tenure weight so each worker $\beta^y_{\Theta\theta}$ reflects the person-weighted average productivity of race-pair $\Theta\theta$ rather than the person-month-weighted average, thus mitigating attrition bias.[23] Based on (3), define

$$\hat{\Delta}^y_{\Theta\Theta'} = \underbrace{(\hat{\beta}^y_{\Theta\theta} - \hat{\beta}^y_{\Theta\theta'})}_{\text{Productivity gap for } \Theta} - \underbrace{(\hat{\beta}^y_{\Theta'\theta} - \hat{\beta}^y_{\Theta'\theta'})}_{\text{Productivity gap for } \Theta'} . \tag{12}$$

Letting Θ be "Hispanic" and Θ' be "white", this compares the productivity boost Hispanic workers get from having a Hispanic manager to the productivity boost that white workers get from having a Hispanic manager. Rewriting (12), we can interpret $\frac{1}{2}\hat{\Delta}^y_{\Theta\Theta'} = \frac{1}{2}(\hat{\beta}^y_{\Theta\theta} + \hat{\beta}^y_{\Theta'\theta'}) - \frac{1}{2}(\hat{\beta}^y_{\Theta\theta'} + \hat{\beta}^y_{\Theta'\theta})$ as the difference between average same-race productivity and average cross-race productivity.

To estimate the variance of productivity of a race-pair, $\sigma^2_{\Theta\theta}$, we use a simplified version of the Kline, Saggio, and Sølvsten (2020) leave-out estimator.[24] Letting e_{it} be i's residual at time t, we first estimate the monthly noise:

$$\hat{\sigma}^2_{\epsilon,\Theta\theta} = \frac{1}{N_{\Theta\theta}} \sum_{i=1}^{N_{\Theta\theta}} \frac{1}{T_i - 1} \sum_{t=1}^{T_i} (e_{it} - \bar{e}_i)^2$$

where $\bar{e}_i := \left(\sum_{t=1}^{T_i} e_{it}\right)/T_i$ is worker i's average residual. We allow $\hat{\sigma}^2_{\epsilon,\Theta\theta}$ to differ at the race-pair level to capture, say, heterogeneous volatility at different stores. We then estimate productivity variance by subtracting the monthly variance from the across-worker variance,[25]

$$\hat{\sigma}^2_{\Theta\theta} = \frac{1}{N_{\Theta\theta}} \sum_{i=1}^{N_{\Theta\theta}} \left((\bar{e}_i - \bar{e}_{\Theta\theta})^2 - \frac{1}{T_i} \hat{\sigma}^2_{\epsilon,\Theta\theta} \right) \tag{13}$$

where $\bar{e}_{\Theta\theta} := \left(\sum_{i=1}^{N_{\Theta\theta}} \bar{e}_i\right)/N_{\Theta\theta}$ is the average residual across workers. Based on (5), define

$$\hat{\Delta}^v_{\Theta\Theta'} = \frac{\hat{\sigma}^2_{\Theta\theta}/\hat{\sigma}^2_{\Theta\theta'}}{\hat{\sigma}^2_{\Theta'\theta}/\hat{\sigma}^2_{\Theta'\theta'}} \tag{14}$$

[22] We loosely refer to this as "inverse-tenure weighting", but T_i may differ from tenure if we have missing observations. For example, some workers are hired before our data starts.

[23] Suppose worker i is employed for six months while worker j is employed for 12 months. The OLS estimate gives of $\beta^y_{\Theta\theta}$ gives worker j twice the weight in the regression as worker i. Since longer tenure workers are positively selected, then $\hat{\beta}^y_{\Theta\theta}$ overestimates the true average productivity. Moreover, we expect this selection will have more effect on cross-race pairs (which have the highest turnover by Proposition 1(d)-3(d)), meaning the OLS estimate of $\hat{\Delta}^y_{\Theta\Theta'}$ will be biased downwards. Weighting by the inverse of tenure ensures each worker has equal weight and avoids this source of bias. WLS may also be more efficient that OLS if the variance of u_i is large relative to the variance of $\epsilon^y_{i,t}$.

[24] Equation (13) coincides with KSS equation (4), but we estimate monthly variance at the race-pair level (rather than the individual level) and weight by person (rather than person-month). We do the former for efficiency, and the latter to mitigate attrition bias. The true KSS estimator also re-estimates the regression (11) after leaving out each observation; this is computationally infeasible for our bootstrap.

[25] Intuitively, this formula comes from $E\left[(\bar{e}_i - \bar{e}_{\Theta\theta})^2\right] = \text{Var}\left[\frac{1}{T_i}\sum_{t=1}^{T_i}\left(u_i + \epsilon^y_{i,t}\right)\right] = \hat{\sigma}^2_{\Theta\theta} + \frac{1}{T_i}\hat{\sigma}^2_{\epsilon,\Theta\theta}$.

Rewriting (12), we can interpret $\sqrt{\hat{\Delta}^v_{\Theta\Theta'}} = \sqrt{\hat{\sigma}^2_{\Theta\theta}\hat{\sigma}^2_{\Theta'\theta'}}/\sqrt{\hat{\sigma}^2_{\Theta\theta'}\hat{\sigma}^2_{\Theta'\theta}}$ as the ratio between (geometric) average same-race variance and (geometric) average cross-race variance. We calculate confidence intervals by 10,000 bootstraps stratified by race pairs.[26]

Proposition 4 shows that we can identify the source of discrimination from $\hat{\Delta}^y_{\Theta\Theta'}$ and $\hat{\Delta}^v_{\Theta\Theta'}$ (see Figure 2). If $\hat{\Delta}^y_{\Theta\Theta'} < 0$ then mean productivity is lower under same-race managers and there must be taste-based discrimination. If $\hat{\Delta}^v_{\Theta\Theta'} < 1$ then productivity variance is lower under same-race managers and there must be screening discrimination. As illustrated in Figure 2, if $\hat{\Delta}^y_{\Theta\Theta'} > 0$ and $\hat{\Delta}^v_{\Theta\Theta'} > 1$ there must be complementary production.

A few remarks are warranted. First, since we are comparing the relative productivity of workers under different races of managers, this is a test of relative discrimination rather than absolute discrimination. Second, in Section 4.2 we discuss the sensitivity of our estimate of the productivity regression (11) to dropping the tenure control, using OLS rather than WLS, and restricting the sample to the first 3 and 6 months of a worker's tenure. Third, we assume that a salesperson's productivity is just determined by their innate ability and, in the case of complementary production, their manager. In Online Appendix B.2 we discuss the possibility that productivity may be correlated with the race of customers or teammates and include a measure of each in the productivity regression (11).

Turnover. Our model predicts that same-race workers have a lower turnover rate no matter the source of discrimination (see Proposition 5(b)). To implement the test, we run the following regression for each race-Θ worker,

$$TURN_i = \beta^x_{\Theta\theta} + \text{controls}_{i,t} + \epsilon^x_i \tag{15}$$

where the indicator $TURN_i$ equals 100 if worker i has left the job in the first six months, and $\beta^x_{\Theta\theta}$ is the turnover of worker Θ under manager θ. We allow for the same controls as in (9). Based on (8), define

$$\hat{\Delta}^x_{\Theta\Theta'} = (\hat{\beta}^x_{\Theta\theta} - \hat{\beta}^x_{\Theta\theta'}) - (\hat{\beta}^x_{\Theta'\theta} - \hat{\beta}^x_{\Theta'\theta'}) \tag{16}$$

Proposition 5(b) predicts that $\hat{\Delta}^x_{\Theta\Theta'} \leq 0$. That is, the white-Hispanic turnover gap is lower under white managers than under Hispanic managers. In Online Appendix B.1, we examine how the results change for cutoffs other than 6 months.

[26] Bootstrapping is useful because $\hat{\Delta}^v_{\Theta\Theta'}$ is a non-standard statistic and because it allows us to perform joint tests with productivity mean and variance. Clustering by worker leads to similar confidence intervals for $\hat{\Delta}^y_{\Theta\Theta'}$.

Table 2: **Hiring tests**

	White worker	Black worker	Hispanic worker
White manager	65.0 [64.6, 65.4]	18.9 [18.6, 19.2]	16.1 [15.8, 16.4]
Black manager	61.7 [60.6, 62.8]	22.0 [21.0, 22.9]	16.3 [15.5, 17.2]
Hispanic manager	60.5 [59.4, 61.6]	19.3 [18.3, 20.2]	20.2 [19.4, 21.1]

Notes: Reports results for 48,755 commissioned salespeople, as described in the text. Below each coefficient is the 90% confidence interval.

4 Results

In this section we apply our tests of relative discrimination. In Section 4.1 we present our hiring results, showing that all races are significantly more likely to hire same-race recruits. In Section 4.2 we examine the results for productivity, where we test for different types of discrimination. In Section 4.3 we consider turnover.

4.1 Hiring tests

We first present our test for discrimination in hiring. This is a test about the existence of discrimination rather than the cause (Proposition 5(a)). Table 2 shows that, for all three races, same-race hiring is higher than cross-race hiring, $\hat{\beta}^r_{\Theta\theta} > \hat{\beta}^r_{\Theta\theta'}$, and this difference is significant at the 1% level for all race pairs. These inequalities imply that hiring is log-supermodular $\hat{\Delta}^r_{\Theta\Theta'} = \hat{\beta}^r_{\Theta\theta}\hat{\beta}^r_{\Theta'\theta'}/\hat{\beta}^r_{\Theta\theta'}\hat{\beta}^r_{\Theta'\theta} > 1$ for each pair of races, consistent with the model.

Quantitatively, the gap between same-race and cross-race hiring $\hat{\beta}^r_{\Theta\theta} - \hat{\beta}^r_{\Theta\theta'}$ is 3.9pp, 2.9pp and 4.0pp for white, black and Hispanic salespeople. The average gap between same-race and cross-race hiring is 3.6pp. This is economically significant. A Hispanic worker is $\hat{\beta}^r_{Hh}/\hat{\beta}^r_{Hw} - 1 = 20.2/16.1 - 1 =$ 25.5% more likely to be hired if the manager is Hispanic than if the manager is white. Given the 50,000 salespeople employed over the course of our study, the firm would employ over 2,000 extra Hispanic employees if managers were all Hispanic as compared to being all white.

4.2 Productivity tests

We now turn to the productivity tests that allow us to separate our three theories (see Proposition 4). The first column of Table 3 shows that mean productivity is supermodular for all three race combinations, but especially for white-Hispanic and black-Hispanic pairs. The standard errors are

Table 3: **Supermodularity statistics**

	Prod. Mean, $\hat{\Delta}^y$	Prod. Variance, $\hat{\Delta}^v$	Turnover, $\hat{\Delta}^x$
White-Black relations	0.0140 $[-0.0117, 0.0383]$	0.953 $[0.799, 1.131]$	-0.0154 $[-0.0436,\ 0.0127]$
White-Hispanic relations	0.0277 $[\ 0.0049, 0.0510]$	0.914 $[0.759, 1.110]$	-0.0354 $[-0.0698, -0.0011]$
Black-Hispanic relations	0.0429 $[-0.0011, 0.0880]$	0.679 $[0.500, 0.956]$	0.0009 $[-0.0589,\ 0.0607]$

Notes: The left column shows the supermodularity of mean productivity (12), the center column shows the supermodularity of productivity variance (14), and the right column shows the supermodularity of turnover (16). For productivity mean and variance, the results come from 48,755 commissioned salespeople. The underlying regression (11) is shown in Table B1 in Online Appendix B. For turnover, the results comes from 39,339 commissioned salespeople. The underlying regression (15) is shown in Table B3 in Online Appendix B. Below each coefficient is the 90% confidence interval calculated using 10,000 bootstraps.

fairly large but the former is significantly different from zero at the 10% level. Quantitatively, the mean productivity gap between same-race and cross-race managers is $\hat{\Delta}^y_{WH}/2 = 1.4$pp for white-Hispanic pairs and $\hat{\Delta}^y_{BH}/2 = 2.1$pp for black-Hispanic pairs. To put this in context, a 1% increase in a store's sales increases its profits by about 6%.[27]

The second column of Table 3 shows that productivity variance is log-submodular for all three race combinations, especially for black-Hispanic pairs. The standard errors are fairly large but the latter is significantly different from one at the 10% level. Quantitatively, the same-race variance is $\sqrt{\hat{\Delta}^v_{BH}} = 0.82$ of the size of cross-race variance for black-Hispanic pairs.

Discrimination tests. Table 4 shows our tests for the source of discrimination, corresponding to the theoretical predictions in Figure 2 (right). Figure 3 illustrates the bootstrapped distributions of $(\hat{\Delta}^y_{\Theta\Theta'}, \hat{\Delta}^v_{\Theta\Theta'})$ corresponding to Figure 2 (left).

The most clear-cut evidence comes from black-Hispanic pairs. Suppose our null hypothesis is "there is no screening discrimination". This translates into

$$H_0 : \Delta^v_{BH} \geq 1 \tag{17}$$

Table 4 shows that 97% of our bootstraps have $\hat{\Delta}^v_{BH} < 1$. As a traditional one-tailed test, one would reject this at the 5% level. This test just concerns the variance, but the supermodular productivity seen in Table 4 indicates there is either complementarity or screening. With this in mind, suppose our null hypothesis is "there is neither screening discrimination nor complementary production."

[27]Suppose the firm's profit function is $\pi = (p-c)q - k$. Of the stores in our sample, the average markup is 31.7%, so $pq = 1.317cq$. Moreover, the average profit margin is 5.11%, so $pq = 1.0511(cq+k)$. Normalizing total costs to 1, we have $pq = 1.0511$, $cq = 0.7470$, $k = 0.2530$, and $\pi = 0.0511$. If sales rise by 1% then we have $pq = 1.0616$, $cq = 0.7545$, $k = 0.2530$, and $\pi = 0.0541$, which is a 5.9% rise in profitability.

Using Figure 2, this translates into

$$H_0 : \Delta^y_{BH} \leq 0 \text{ and } \Delta^v_{BH} \geq 1 \tag{18}$$

We can reject this hypothesis since 99.6% of our bootstraps have either $\hat{\Delta}^y_{BH} > 0$ or $\hat{\Delta}^v_{BH} < 1$.[28]

For white-Hispanic pairs only 77% of our bootstraps exhibit $\hat{\Delta}^v_{BH} < 1$, so we cannot conclude there must be screening discrimination (17). However, the supermodular productivity means that 99.1% of our bootstraps have either $\hat{\Delta}^y_{BH} > 0$ or $\hat{\Delta}^v_{BH} < 1$ so we can reject the null hypothesis that there is neither screening discrimination nor complementary production (18).

For white-black pairs, there is little evidence in favor of any one hypothesis. This does not mean that there is no discrimination; recall from Table 2 that white and black managers hire relatively more of their own race. Rather, all three types of discrimination are likely present, with no one force outweighing the others. This follows from the fact that the three vectors $(\partial_z, \partial_p, \partial_k)$ span the $(\Delta^y_{\Theta\Theta'}, \Delta^v_{\Theta\Theta'})$ space in Figure 2 (left). Thus, if $(\Delta^y_{BW}, \Delta^v_{BW}) = (0, 1)$ and there is non-trivial discrimination so $(d, z, k) \neq (0, 0, 0)$ then all three forces must be present, $(d, z, k) \gg (0, 0, 0)$.[29]

Overall, the hiring results in Table 2 show us that there is discrimination with all three pairs of races. Table 4 tells us there is screening discrimination for black-Hispanic pairs, and some combination of screening discrimination and complementary production for white-Hispanic pairs. We note some caveats when interpreting the results. First, we cannot rule out taste-discrimination for any pair. It may be that different types of discrimination coexist and offset one another. Second, we do not know which manager races are driving the supermodularity. If productivity variance is log-submodular, Proposition 4(b) tells us that $z_{\Theta\theta'} + z_{\Theta'\theta} > 0$, but this does not tell us whether the screening discrimination is coming from θ managers, θ' managers or both. Third, this is a test of relative discrimination; if managers of all races discriminate against one race, we will not detect it.

Quantile analysis. We can get further insight by looking at the supermodularity of output quantiles. This helps us visualize the data and better understand our supermodularity statistics in Table 3. For example, for black-Hispanic pairs, is lower same-race variance driven by fewer extreme outcomes at the top of the distribution, the bottom, or both? Propositions 1-3 do not directly address quantiles, but the right-hand side of Figure 1 illustrates the qualitative predictions of the three theories.

Let $y^q_{\Theta\theta}$ be the q^{th} quantile of output for worker Θ under manager θ, and define $\Delta^q_{\Theta\Theta'} :=$

[28] This is a two-dimensional test which makes the interpretation of rejection unusual. If $\Delta^y_{BH} = 0$ and $\Delta^v_{BH} = 1$ one would expect 25% of the data to be in the top-left quadrant ($\hat{\Delta}^y_{BH} \leq 0$ and $\hat{\Delta}^v_{BH} \geq 1$). In comparison if $\Delta^v_{BH} = 1$ in a one-dimensional test like (17) then we would expect to see 50% of the data in the top two quadrants ($\hat{\Delta}^v_{BH} \geq 1$). In our data, we see that 0.4% of the bootstraps are in the top-left quadrant, which is thus equivalent to 0.8% in one-dimensional test.

[29] More specifically, (d, z, k) solve $(d, z, k)\big(\partial_p, \partial_z, \partial_k\big)' = \big(0, 0\big)$ where ∂_p is a column vector and $d = d_{Wb} + d_{Bw}$, $z = z_{Wb} + z_{Bw}$ and $k = k_{Wb} + k_{Bw}$. For example, if $\partial_p = (-1, 1/2)'$, $\partial_z = (1/2, -1)'$ and $\partial_k = (1, 1)'$ then $z = d = 2k$.

Table 4: **Tests of discrimination**

	$\hat{\Delta}^y_{WB} < 0$	$\hat{\Delta}^y_{WB} > 0$
$\hat{\Delta}^v_{WB} > 1$	7.8%	23.6%
$\hat{\Delta}^v_{WB} < 1$	10.7%	57.9%

(a) White-Black Relations

	$\hat{\Delta}^y_{WH} < 0$	$\hat{\Delta}^y_{WH} > 0$
$\hat{\Delta}^v_{WH} > 1$	0.9%	22.1%
$\hat{\Delta}^v_{WH} < 1$	1.4%	75.6%

(b) White-Hispanic Relations

	$\hat{\Delta}^y_{BH} < 0$	$\hat{\Delta}^y_{BH} > 0$
$\hat{\Delta}^v_{BH} > 1$	0.4%	2.6%
$\hat{\Delta}^v_{BH} < 1$	5.2%	91.8%

(c) Black-Hispanic Relations

Notes: These tables show the test of discrimination using the joint distribution of supermodularity statistics $(\hat{\Delta}^y_{\Theta\Theta'}, \hat{\Delta}^v_{\Theta\Theta'})$, as described in Figure 2. These tables are based on 10,000 bootstraps of 48,755 commissioned salespeople.

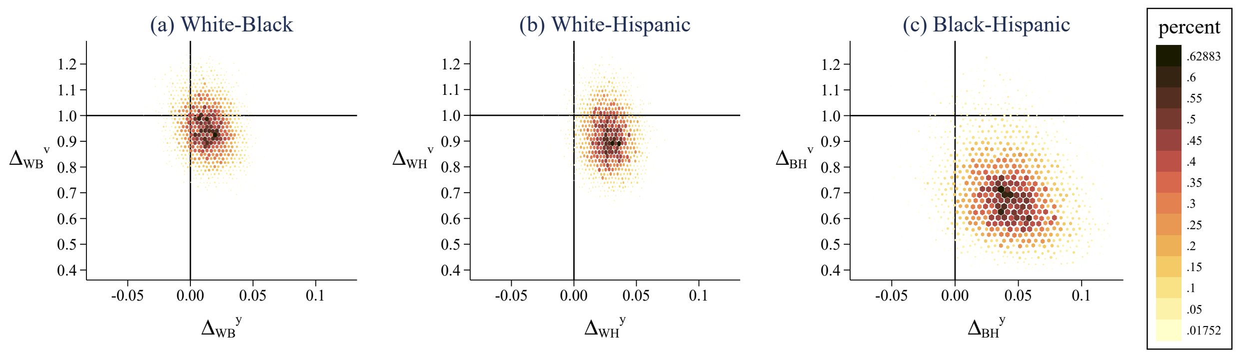


Figure 3: **Distribution of supermodularity statistics**
Notes: This figure shows the joint distribution of the supermodularity statistics $(\hat{\Delta}^y_{\Theta\Theta'}, \hat{\Delta}^v_{\Theta\Theta'})$. This is the empirical counterpart Figure 2. The figure is based on 10,000 bootstraps of 48,755 commissioned salespeople; 99.5% of bootstrap samples fall within the ranges displayed here.

$y^q_{\Theta\theta} + y^q_{\Theta'\theta'} - y^q_{\Theta\theta'} - y^q_{\Theta\theta'}$ as the supermodularity for the q^{th} quantile of output, where $\Theta = \theta$ and $\Theta' = \theta'$. As in Section 2.4, a Taylor approximation implies

$$\Delta^q_{\Theta\Theta'} = -(d_{\Theta\theta'} + d_{\Theta'\theta})\frac{\partial y^q(\mathbf{0})}{\partial d} - (z_{\Theta\theta'} + z_{\Theta'\theta})\frac{\partial y^q(\mathbf{0})}{\partial z} - (k_{\Theta\theta'} + k_{\Theta'\theta})\frac{\partial y^q(\mathbf{0})}{\partial k}$$

From Figure 1, $\partial y^q/\partial d > 0$, $\partial y^q/\partial k < 0$ and $\partial y^q/\partial z < 0$ for most quantiles. To find the empirical equivalent, recall that $\bar{e}_i$ is the within-worker mean residual from the productivity regression (11). Given a quantile $q \in [0, 1]$, let $\bar{e}^q_{\Theta\theta}$ be the residual of the q^{th} highest Θ worker under a θ manager. By construction, $\bar{e}_i$ has zero mean within each race-pair $\Theta\theta$, so $\bar{\varepsilon}^q_{\Theta\theta} := \bar{e}^q_{\Theta\theta} + \hat{\beta}^y_{\Theta\theta}$ is the absolute productivity of the q^{th} highest Θ worker under a θ manager. Then define

$$\hat{\Delta}^q_{\Theta\Theta'} = \bar{\varepsilon}^q_{\Theta\theta} + \bar{\varepsilon}^q_{\Theta'\theta'} - \bar{\varepsilon}^q_{\Theta\theta'} - \bar{\varepsilon}^q_{\Theta'\theta}, \tag{19}$$

where $\Theta = \theta$ and $\Theta' = \theta'$.[30] As in Figure 1, one thus expects $\hat{\Delta}^q_{\Theta\Theta'}$ to be (i) negative and increasing

[30] One concern is that $\bar{\varepsilon}_i$ measures u_i with noise. If the noise only depends on the quantile (e.g. better workers stay

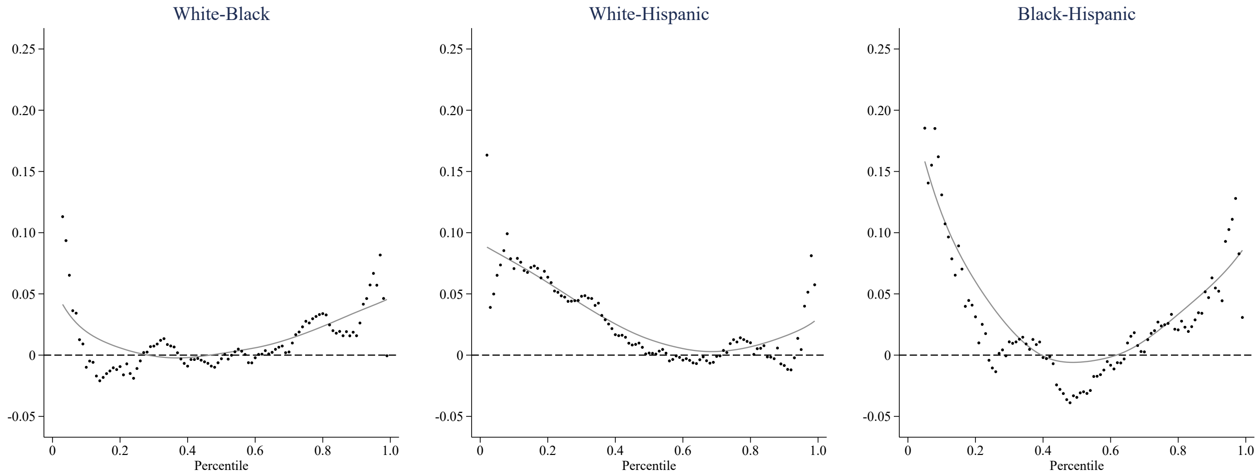


Figure 4: **Supermodularity of productivity by quantile**

Notes: This figure shows $\hat{\Delta}^q_{\Theta\Theta'}$ as defined in (19) across 98 one-percentile bins, after we have thrown away the top and bottom bins. This is the empirical counterpart to Figure 1 (right). The black-Hispanic figure omits the lowest data (value 0.66) point that does not fit in the figure. The Lowess curve has bandwidth 0.8.

under taste-discrimination, (ii) positive and decreasing under screening discrimination, and (iii) positive and increasing under complementary production.

Figure 4 shows the results. The three pictures are similar, showing a "ski jump" style curve. At the bottom of the distribution, $\hat{\Delta}^q_{\Theta\Theta'}$ is initially positive and decreasing, consistent with screening discrimination. This is particularly stark with black-Hispanic pairs: For example, at the 10th percentile the difference between same-race and cross race productivity $\hat{\Delta}^q_{BH}/2 = 0.065$ (or 6.8%). Around the median worker, there is little difference between same-race and cross-race workers. At the top, the gap becomes positive again, perhaps indicative of some complementarity. (In the model, we assume that complementarity is uniform across all workers, but this need not be the case.) These pictures are similar if we run (11) with no tenure control or with no controls at all.

Sensitivity. Table B2 in Online Appendix B shows how the specification of the productivity regression (11) affects the results in Tables 3 and 4. Column (2) drops the tenure control; this has little effect on the estimates. Columns (3) and (4) restrict the sample to workers in their first three and six months, respectively. Mean supermodularity $\Delta^y_{\Theta\Theta'}$ is broadly similar, while the submodularity of the variance $\Delta^v_{\Theta\Theta'}$ is muted. In unreported regressions, we explore the trajectory of these coefficients; when we truncate at 12 months the results are similar to the baseline results. Columns (5)-(6) use OLS with and without the tenure control. Mean supermodularity $\Delta^y_{\Theta\Theta'}$ is

longer and have less noise) then this should cancel out in the supermodularity calculation and should not bias $\Delta^q_{\Theta\Theta'}$. However, if same-race workers have substantially longer tenure (and thus less noise) than cross-race workers, $\Delta^q_{\Theta\Theta'}$ would tend pivot clockwise around its mid-point. The differences in tenure is small, so we do not expect this effect to be important in practice.

smaller while the variance $\Delta^v_{\Theta\Theta'}$ is similar to the baseline; the former may be due to the downward bias discussed in footnote 23. Columns (7)-(8) show the OLS estimates of $\Delta^y_{\Theta\Theta'}$ increase when we truncate the samples at three and six months. Looking across these regressions, there is consistent evidence that black-Hispanic pairs exhibit supermodular mean productivity and log-submodular productivity variance, indicative of screening or complementarity. Figure B2 in Online Appendix B also recalculates Figure 4 for (a) the first six months of workers' tenure, and (b) later months. The "ski-jump" pattern is consistent across both periods.

4.3 Turnover results

Finally, we come to turnover. All three models predict that turnover is submodular. The third column of Table 3 shows that $\hat{\Delta}^q_{\Theta\Theta'}$ is negative for two pairs and essentially zero for the third; the white-Hispanic combination is significantly negative at the 10% level. Thus we do not reject the model. Quantitatively, the average difference between same-race and cross-race turnover rates is $(\hat{\Delta}^q_{WB} + \hat{\Delta}^q_{WH} + \hat{\Delta}^q_{BH})/6 = -0.8\%$ compared to a base six-month turnover rate of 36.1%. We hesitate to read to much into these results for the reasons discussed in Section 2.5.

5 Discussion

This section discusses how one might identify the bias of an individual race from the three pair-wise tests, and shows how symmetric discrimination in hiring leads to the under-representation of minorities. In Online Appendix B we examine the sensitivity of the empirical results to different specifications and discuss how alternative theories (e.g. customer bias, peer effects, favoritism) fit into our framework.

5.1 Triangulating individual bias

Our supermodularity test examines relative bias. This has the advantage that it allows for worker and manager heterogeneity, and overcomes the problem of infra-marginality in the outcomes test. The downside is that it does not identify the bias of any one type of manager. For example, the fact that $\Delta^y_{BH} > 0$ could be driven by black managers, Hispanic managers or both. However, looking at Table 3, the three mean supermodularity statistics are $(\Delta^y_{WB}, \Delta^y_{WH}, \Delta^y_{BH}) = (0.0140, 0.0277, 0.0429)$, which suggests that Hispanics play a large role in driving these results. We now propose a formalization of this intuition.

Say that managers *discriminate symmetrically* if

$$d_{\Theta'\theta} = d_{\Theta''\theta} =: d_\theta, \; z_{\Theta'\theta} = z_{\Theta''\theta} =: z_\theta, \text{ and } k_{\Theta'\theta} = k_{\Theta''\theta} =: k_\theta, \tag{20}$$

so that manager θ treats all cross-race workers symmetrically. Assumption (20) makes sense if managers discriminate in favor of same-race candidates rather than against people of other races (e.g. because of shared culture). Empirically, this is consistent with our hiring results in Table 2, since we cannot reject that cross-race hiring rates are statistically different from one another.[31]

Under the symmetry assumption (20), the supermodularity statistic (3) becomes

$$\Delta^y_{\Theta\Theta'} = -(d_\theta + d_{\theta'})\frac{\partial \bar{y}(\mathbf{0})}{\partial d} - (z_\theta + z_{\theta'})\frac{\partial \bar{y}(\mathbf{0})}{\partial z} - (k_\theta + k_{\theta'})\frac{\partial \bar{y}(\mathbf{0})}{\partial k}.$$

Using the three supermodularity estimates, we can then triangulate the impact of manager θ on the mean productivity of cross-race workers,

$$\Delta^y_\theta := \frac{1}{2}\left(\Delta^y_{\Theta\Theta'} + \Delta^y_{\Theta\Theta''} - \Delta^y_{\Theta'\Theta''}\right) = -d_\theta\frac{\partial \bar{y}(\mathbf{0})}{\partial d} - z_\theta\frac{\partial \bar{y}(\mathbf{0})}{\partial d} - k_\theta\frac{\partial \bar{y}(\mathbf{0})}{\partial d}$$

Similarly, we can derive the impact of manager θ on the variance,

$$\log \Delta^v_\theta := \frac{1}{2}\left(\log \Delta^v_{\Theta\Theta'} + \log \Delta^v_{\Theta\Theta''} - \log \Delta^v_{\Theta'\Theta''}\right) = -d_\theta\frac{\partial \nu(\mathbf{0})}{\partial d} - z_\theta\frac{\partial \nu(\mathbf{0})}{\partial d} - k_\theta\frac{\partial \nu(\mathbf{0})}{\partial d}$$

We can then apply Proposition 4 to test for bias of individual managers. In particular, if $\Delta^y_\theta < 0$ then manager θ exhibits taste-based bias, if $\Delta^v_\theta < 1$ then manager θ exhibits screening discrimination, and if $\Delta^y_\theta > 0$ and $\Delta^v_\theta > 1$ then manager θ exhibits complementary production.

We illustrate these ideas with the numbers in Table 3. For productivity mean, $(\hat{\Delta}^y_w, \hat{\Delta}^y_b, \hat{\Delta}^y_h) = (-0.0006, 0.0146, 0.0283)$, and for the variance, $(\hat{\Delta}^v_w, \hat{\Delta}^v_b, \hat{\Delta}^v_h) = (1.133, 0.841, 0.807)$. Applying the supermodularity tests in Proposition 4, black and Hispanic managers have supermodular productivity and submodular variance, suggestive of screening discrimination, whereas white managers have modular productivity and supermodular variance, suggestive of a mix of taste-based discrimination and complementarity. These results rely on the symmetry assumption (20), so one should interpret them with caution. But this approach illustrates the benefit of studying multiple racial groups.

5.2 Long-run implications of discrimination

The results presented in Section 4 provide evidence of screening discrimination, but not of taste-based discrimination. This conclusion may be somewhat of a relief. It is not a surprise that people with similar backgrounds see past stereotypes and are better at assessing individual characteristics. And this feels less pernicious than if managers have extensive taste-based biases. However, we now argue that screening discrimination systematically disadvantages minority candidates.[32]

Consider a simple model of employment dynamics based on Section 2. There is a continuum of

[31] Formally, the hypothesis that $r_{Wb} = r_{Wh}$ has a p-value of 0.23, $r_{Bw} = r_{Bh}$ has a p-value of 0.55, and $r_{Hw} = r_{Hb}$ has a p-value of 0.69. Assumption (20) is actually weaker than this, e.g. if white managers hire a lot of whites then $r_{Bw} < r_{Bh}$, even if white managers treat black and Hispanic workers symmetrically.

[32] This logic is similar to Müller-Itten and Öry (2022), which focuses on mentoring.

agents who belong to one of two races. Fraction $n > \frac{1}{2}$ of the population is of the "majority" race and fraction $1 - n$ is of the "minority" race. The races are otherwise symmetric. A firm initially employs fraction M_0 managers and N_0 workers of the majority race. These fractions $\{M_t, N_t\}$ evolve over continuous time. At each point in time, workers and managers quit at an exogenous rate and are replaced. When a worker leaves, a random manager sequentially considers applicants until a worker's signal passes the manager's hiring threshold. As in Propositions 1(a)-3(a), a manager hires each same-race candidate with probability p_s and each cross-race candidate with probability $p_c < p_s$; by symmetry, these numbers are the same for majority and minority managers. When a manager leaves, a random worker is promoted into the vacancy.

Proposition 6. *Starting from any initial condition $\{M_0, N_0\}$, the system converges to a unique steady state $\{M^*, N^*\}$. Workers and managers have the same composition of races, and the majority population is over-represented, $M^* = N^* > n$.*

Proof. See Online Appendix C.2. $\square$

Minorities are at a disadvantage simply because they are minorities. Intuitively, the problem is that most minorities are interviewed by managers who discriminate against them; the exact source of the discrimination does not matter. We can use the hiring numbers in Table 2 to understand the quantitative significance of the result. Given these hiring numbers, the firm converges to a steady state of $(N_W^*, N_B^*, N_H^*) = (63.6, 19.6, 16.8)$.[33] We don't know what "neutral hiring" looks like, since we do not know the bias of the different managers, but if we pick a manager race $\theta \in \{w, g, h\}$ with probability $\{1/3, 1/3, 1/3\}$, we obtain $(N_W, N_B, N_H) = (62.4, 20.1, 17.5)$ by taking the average of the columns in Table 2. Thus white workers are over-represented by 1.2pp while Hispanics are under-represented by 0.7pp. This is quantitatively significant: There are 4.2% fewer Hispanics employed than in a "neutral" workplace simply because the Hispanics are in the minority. Moreover this calculation may underestimate the true effect: If productivity is supermodular in race (as in Table 3), minority workers and managers have lower average productivity than their majority counterparts; one would thus expect minority workers to be promoted to managerial positions less frequently than majority workers.

This "institutional bias" can be mitigated if the firm promotes more minority managers. This is especially useful as a way to get to steady state, but it is harder to use such a policy to shift an inequitable steady state. Indeed, Appendix C.2 shows that as the number of minority workers vanish, the number of minority managers would have to equal $p_c/(p_c+p_s)$ to achieve equitable recruitment, meaning the ratio of minority managers to workers grows without bound. A different strategy is to centralize recruitment and take away the job of hiring from the day-to-day manager. For example,

[33] Denote the matrix in Table 2 by r, where the columns represent the fraction of worker Θ hired by the various manager races. Then the steady state is given by the fixed point, $(N_W^*, N_B^*, N_H^*) = (N_W^*, N_B^*, N_H^*)r$.

one scheme would match the candidate with the interviewer best equipped to interview them on a variety of characteristics (e.g. race, age, school, gender). One can then evaluate the interviewer based on their acceptance rate and the subsequent productivity of their hires.

6 Conclusion

This paper proposes and implements a method to separate three theories of relative discrimination (taste-based discrimination, screening discrimination, complementary production). We derive the predictions of the theories in a common framework, provide a test to identify the type of discrimination from any data that feature manager and worker characteristics as well as productivity, and examine the implications for a firm with 48,755 salespeople. In our setting, we find statistically and economically significant discrimination in hiring for all three pairs of races. For black-Hispanic pairs, screening discrimination is dominant; for white-Hispanic pairs, our test indicates a combination of screening discrimination and complementary production; for white-black pairs, all three types of discrimination seem to play a role. The results for black-Hispanic pairs, including high cross-race productivity variance, are especially notable and could be due to a number of factors. Perhaps the simplest explanation could relate to Lazear's (1999) model of cultural assimilation, whereby minorities have a greater incentive to assimilate with the majority culture rather than other minorities. In this case, black and Hispanic hiring managers may be better able to assess white candidates than each other, although testing such a hypothesis would require a measure of acculturation. More broadly, our finding for black-Hispanic pairs highlights the importance of studying how minority groups interact with each other.

Our tests of discrimination can be applied by firms to measure discrimination along the lines of race, gender, language, age, disability, educational background, sexual orientation, and national origin. For instance, companies can use real-time performance data to see whether managers appear to have a taste-based bias against certain groups, whether they are poorly screening different groups, or are less effective at managing them. Then, depending on their results, they may decide whether to focus diversity efforts on quotas and debiasing managers, using algorithmic tools or diverse hiring committees to assist them in the screening process, or providing training to enhance their ability to mentor those groups when they arrive. These results can also inform how firms assemble hiring committees: should it send alumni back to their own college, which may enhance screening but come at the cost of taste-based biases? The method can also be generalized to understand bias outside the employment setting, as long as a continuous outcome variable is available. For instance, the method could be used to assess bias in college admissions (from GPA distributions) or bias in VC funding (from firm value growth). We anticipate that the growth of high-resolution productivity data in many walks of life will present several opportunities and applications.

Appendix

A Proofs for Section 2

A.1 Useful properties of the normal distribution

Normal distribution. Let $\phi(\alpha) = \exp(-\alpha^2/2)/\sqrt{2\pi}$ be the density of the normal distribution, and Φ be the corresponding cdf. We note the derivative is

$$\phi'(\alpha) = -\alpha\phi(\alpha). \tag{21}$$

Moreover, the density is log-concave since $(\log\phi)'' = -1$.

Truncated normal. Given a normal random variable $\hat{y} \sim N(\mu - k, \eta^2)$ and cutoff $y^* + d$, the expectation and variance of the truncated normal are given by,[34]

$$E[\hat{y}|\hat{y} \geq y^* + d] = \mu - k + \lambda(\alpha)\eta \tag{22}$$

$$Var[\hat{y}|\hat{y} \geq y^* + d] = \eta^2(1 + \alpha\lambda(\alpha) - \lambda(\alpha)^2) \tag{23}$$

where $\alpha := (y^* + d - (\mu - k))/\eta$ is the normalized cutoff and $\lambda(\alpha) = \phi(\alpha)/(1 - \Phi(\alpha))$ is the hazard rate of the standard normal. We recall the estimator variance $\eta = \sqrt{\sigma_0^4/(\sigma_0^2 + \sigma_\epsilon^2(1+z)^2)}$ and write $\eta' = \partial\eta/\partial z < 0$; greater signal noise reduces the estimator variance, as discussed in Section 2. The normalized cutoff then depends on our parameters as follows

$$\frac{\partial\alpha}{\partial d} = \frac{1}{\eta} > 0 \qquad \frac{\partial\alpha}{\partial z} = -\frac{\alpha\eta'}{\eta} > 0 \qquad \frac{\partial\alpha}{\partial k} = \frac{1}{\eta} > 0. \tag{24}$$

Write $\delta = \hat{y} - (y^* + d) \geq 0$ for the *excess productivity*, i.e. the random variable that measures the amount by which expected productivity $\hat{y}$ exceeds the hiring threshold $y^* + d$.

Lemma 1. *Excess productivity δ falls in d, z, and k in the likelihood ratio order (MLRP).*

Proof. The density of δ is given by

$$g(\delta) = g(\delta; d, z, k) := \frac{1}{\eta(z)}\phi\left(\frac{y^* + d + \delta - (\mu - k)}{\eta(z)}\right) \Big/ \left(1 - \Phi\left(\frac{y^* + d + \delta - (\mu - k)}{\eta(z)}\right)\right).$$

Thus, δ MLRP-decreases in d if

$$\frac{g(\bar{\delta}; \bar{d}, z, k)}{g(\delta; \bar{d}, z, k)} \leq \frac{g(\bar{\delta}; d, z, k)}{g(\delta; d, z, k)}$$

for all $\delta < \bar{\delta}$ and $d \leq \bar{d}$. Rearranging, we obtain

$$\log g(\bar{\delta}; \bar{d}, z, k) + \log g(\delta; d, z, k) \leq \log g(\bar{\delta}; d, z, k) + \log g(\delta; \bar{d}, z, k),$$

[34] Derived with integration by parts, or found on, e.g., https://en.wikipedia.org/wiki/Truncated_normal_distribution

which is to say that the density is log-submodular in δ and d. Differentiating,

$$\frac{\partial^2(\log g)}{\partial y^*\partial\delta}=\frac{\partial^2}{\partial d\partial\delta}\log\phi\left(\frac{y^*+d+\delta-(\mu-k)}{\eta(z)}\right)=\frac{1}{\eta(z)^2}\left(\log\phi\left(\frac{y^*+d+\delta-(\mu-k)}{\eta(z)}\right)\right)''<0$$

as required. This argument also shows that δ MLRP-increases in k.

To see that δ MLRP-decreases in z, we wish to show that $g(\delta; d, z, k)$ is log-submodular in δ and z. We write $\psi(\delta, z)=\frac{y^*+d+\delta-(\mu-k)}{\eta(z)}$, note the derivatives $\frac{\partial\psi}{\partial\delta}>0$, $\frac{\partial\psi}{\partial z}>0$, $\frac{\partial^2\psi}{\partial\delta\partial z}>0$, and compute

$$\frac{\partial^2(\log g)}{\partial\delta\partial z}=\frac{\partial^2}{\partial\delta\partial z}\log\phi(\psi(\delta,z))=\frac{\partial\psi}{\partial\delta}\cdot\frac{\partial\psi}{\partial z}\cdot[\log\phi(\psi(\delta,z))]''+\frac{\partial^2\psi}{\partial\delta\partial z}\cdot[\log\phi(\psi(\delta,z))]'<0.$$

□

Lemma 2. *The hazard rate of the standard normal $\lambda(\alpha)$ has the following properties:*
(a) $\lambda(\alpha)$ is increasing with derivative $\lambda'=\lambda(\lambda-\alpha)\in(0,1)$,
(b) $\lambda(\alpha)$ is bounded below by $\lambda(\alpha)>\alpha>\alpha\lambda'(\alpha)$,
(c) $\lambda(\alpha)$ is convex with second derivative $\lambda''=(2\lambda-\alpha)\lambda'-\lambda$.

Proof. We prove (a) and (b) out of order. Using (22), first note that $\lambda(\alpha)=E[\zeta|\zeta\geq\alpha]>\alpha$, where $\zeta\sim N(0,1)$. Using (21), this has slope $\lambda'=\lambda(\lambda-\alpha)$, which is positive given $\lambda>\alpha$. To prove the slope is less that one, we specialize Lemma 1 to the standard normal with cutoff α, and excess $\delta=\zeta-\alpha$ with density $g(\delta;\alpha)$, so that

$$\lambda(\alpha)=E[\zeta|\zeta\geq\alpha]=\alpha+\int_{\delta=0}^{\infty}\delta g(\delta;\alpha)d\delta. \tag{25}$$

By the proof of Lemma 1, δ MLRP-falls in α. Hence the second term in (25) falls in α, so $\lambda'(\alpha)<1$. This completes the proof of (a) and (b). The convexity of the hazard rate is well known.[35] □

A.2 Proof of Propositions 1-4

Propositions 1-3 follow immediately from (2), which we replicate here for convenience

$$\begin{array}{rllll}
\text{Taste-Based Discrimination:} & \frac{\partial p}{\partial d}<0, & \frac{\partial\bar{y}}{\partial d}>0, & \frac{\partial v}{\partial d}<0, & \frac{\partial x}{\partial d}>0\\
\text{Screening Discrimination:} & \frac{\partial p}{\partial z}<0, & \frac{\partial\bar{y}}{\partial z}<0, & \frac{\partial v}{\partial z}>0, & \frac{\partial x}{\partial z}>0\\
\text{Complementary Production:} & \frac{\partial p}{\partial k}<0, & \frac{\partial\bar{y}}{\partial k}<0, & \frac{\partial v}{\partial k}<0, & \frac{\partial x}{\partial k}>0
\end{array}$$

While the propositions are organized by theory (the rows in the matrix), it is convenient to organize the proof by the outcome variables (i.e. the columns of the matrix).

[35] E.g., https://math.stackexchange.com/questions/1349555/standard-normal-distribution-hazard-rate

Hiring probability. The hiring probability $p = \Pr(\hat{y} > y^* + d) = 1 - \Phi(\alpha)$ clearly falls in α and thus, by (24), in d, z, and k.[36]

Turnover. Recalling the excess productivity $\delta := \hat{y} - (y^* + d) \geq 0$ from Lemma 1, and writing $\tilde{\epsilon} := \hat{y} - y \sim N(0, \gamma^2)$ for the difference between estimated and actual productivity, the worker is fired iff $\tilde{\epsilon}$ exceeds the sum of the two safety margins δ and τ. Conditional turnover "$x|\delta$" thus equals $\Pr(\tilde{\epsilon} \geq \delta + \tau) = 1 - \Phi((\delta + \tau)/\gamma(z))$, which falls in δ and rises in the residual variance $\gamma^2(z) = \sigma_0^2 \sigma_\epsilon^2 (1+z)^2 / (\sigma_0^2 + \sigma_\epsilon^2 (1+z)^2)$, which in turn rises in z. In expectation over δ, turnover becomes

$$x = \Pr(\tilde{\epsilon} \geq \delta + \tau | \delta \geq 0) = \int_{\delta \geq 0} (1 - \Phi((\delta + \tau)/\gamma(z))) g(\delta; d, z, k) d\delta. \tag{26}$$

By Lemma 1, a rise in d or k MLRP-decreases excess productivity δ and so MLRP-raises conditional turnover $x|\delta$. A rise in z also MLRP-raises $x|\delta$, and additionally raises $x|\delta = 1 - \Phi((\delta + \tau)/\gamma(z))$ directly, so turnover rises.

Expected productivity. Differentiating (22) $\bar{y} = E[\hat{y}|\hat{y} \geq y^*] = \mu - k + \lambda(\alpha)\eta$

$$\frac{\partial \bar{y}}{\partial d} = \lambda' > 0 \qquad \frac{\partial \bar{y}}{\partial z} = \left(\lambda - \alpha\lambda'\right)\eta' < 0 \qquad \frac{\partial \bar{y}}{\partial k} = -1 + \lambda' < 0, \tag{27}$$

where the derivatives rely on (24), and the inequalities on Lemma 2.

Productivity variance. Applying the law of total variance, using (23), and Lemma 2(a), productivity variance equals

$$\begin{aligned} v = Var(y|\hat{y} \geq y^*) &= E[Var(y|\hat{y})|\hat{y} \geq y^*] + Var(E[y|\hat{y}]|\hat{y} \geq y^*) \\ &= Var(y|\hat{y}) + Var(\hat{y}|\hat{y} \geq y^*) = \gamma^2 + \eta^2(1 + \alpha\lambda(\alpha) - \lambda(\alpha)^2) = \sigma_0^2 - \eta^2\lambda'(\alpha). \end{aligned} \tag{28}$$

Differentiating and using (24) and Lemma 2(b,c)

$$\frac{\partial v}{\partial d} = -\eta\lambda'' = -\eta((2\lambda - \alpha)\lambda' - \lambda) < 0 \qquad \frac{\partial v}{\partial z} = \eta\,(2\lambda - \alpha)\left(\alpha\lambda' - \lambda\right)\eta' > 0 \qquad \frac{\partial v}{\partial k} = \frac{\partial v}{\partial d} < 0 \tag{29}$$

Proof of Proposition 4. Dividing (29) by (27), we get

$$0 < \frac{-\partial v/\partial d}{\partial \bar{y}/\partial d} = \eta[(2\lambda - \alpha) - \lambda/\lambda'] < \eta(2\lambda - \alpha) = \frac{\partial v/\partial z}{-\partial \bar{y}/\partial z}. \tag{30}$$

[36]Here we use the assumption that managers hire less than half of the applicants, $y^* > \mu$, which guarantees $\alpha = y^* + d - (\mu - k) > 0$, which in turn underlies $\partial\alpha/\partial z > 0$ in (24).

References

AGAN, A., AND S. STARR (2017): "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *Quarterly Journal of Economics*, 133(1), 191–235.

ALLPORT, G. W., K. CLARK, AND T. PETTIGREW (1954): "The Nature of Prejudice," .

ALTONJI, J. G., AND R. M. BLANK (1999): "Race and Gender in the Labor Market," *Handbook of Labor Economics*, 3, 3143–3259.

ALTONJI, J. G., AND C. R. PIERRET (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116(1), 313–350.

AMIR, Y. (1969): "Contact Hypothesis in Ethnic Relations.," *Psychological Bulletin*, 71(5), 319.

ANWAR, S., AND H. FANG (2006): "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence," *American Economic Review*, 96(1), 127–151.

——— (2015): "Testing for Racial Prejudice in the Parole Board Release Process: Theory and Evidence," *Journal of Legal Studies*, 44(1), 1–37.

ARNOLD, D., W. DOBBIE, AND P. HULL (2022): "Measuring Racial Bias in Bail Decisions," *American Economic Review*, forthcoming.

ARNOLD, D., W. DOBBIE, AND C. YANG (2018): "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, 133(4), 1885–1932.

ÅSLUND, O., L. HENSVIK, AND O. N. SKANS (2014): "Seeking Similarity: How Immigrants and Natives Manage in the Labor Market," *Journal of Labor Economics*, 32(3), 405–441.

AYRES, I., AND P. SIEGELMAN (1995): "Race and Gender Discrimination in Bargaining for a New Car," *American Economic Review*, pp. 304–321.

BANDIERA, O., I. BARANKAY, AND I. RASUL (2009): "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," *Econometrica*, 77(4), 1047–1094.

BEAMAN, L., AND J. MAGRUDER (2012): "Who gets the Job Referral? Evidence from a Social Networks Experiment," *American Economic Review*, 102(7), 3574–3593.

BECKER, G. S. (1957): *The Economics of Discrimination*. University of Chicago Press.

BENSON, A., AND L.-P. LEPAGE (2022): "Learning to Discriminate on the Job," Working paper, University of Minnesota.

BERTRAND, M., AND S. MULLAINATHAN (2004): "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 94(4), 991–1013.

BLACK, D. A. (1995): "Discrimination in an Equilibrium Search Model," *Journal of Labor Economics*, 13(2), 309–334.

Brown, M., E. Setren, and G. Topa (2016): "Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System," *Journal of Labor Economics*, 34(1), 161–209.

Cornell, B., and I. Welch (1996): "Culture, Information, and Screening Discrimination," *Journal of Political Economy*, 104(3), 542–571.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016): "Referral-Based Job Search Networks," *Review of Economic Studies*, 83(2), 514–546.

Fisman, R., D. Paravisini, and V. Vig (2017): "Cultural Proximity and Loan Outcomes," *American Economic Review*, 107(2), 457–492.

Giuliano, L., D. I. Levine, and J. Leonard (2009): "Manager Race and the Race of New Hires," *Journal of Labor Economics*, 27(4), 589–631.

——— (2011): "Racial Bias in the Manager-Employee Relationship An Analysis of Quits, Dismissals, and Promotions at a Large Retail Firm," *Journal of Human Resources*, 46(1), 26–52.

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

Hedegaard, M. S., and J.-R. Tyran (2018): "The Price of Prejudice," *American Economic Journal: Applied Economics*, 10(1), 40–63.

Hjort, J. (2014): "Ethnic Divisions and Production in Firms," *Quarterly Journal of Economics*, 129(4), 1899–1946.

Kline, P., E. K. Rose, and C. R. Walters (2022): "Systemic Discrimination Among Large U.S. Employers*," *The Quarterly Journal of Economics*, 137(4), 1963–2036.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out Estimation of Variance Components," *Econometrica*, 88(5), 1859–1898.

Knowles, J., N. Persico, and P. Todd (2001): "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 109(1), 203–229.

Lang, K. (1986): "A Language Theory of Discrimination," *Quarterly Journal of Economics*, 101(2), 363–382.

Lang, K., and J.-Y. Lehmann (2012): "Racial Discrimination in the Labor Market: Theory and Empirics," *Journal of Economic Literature*, 50(4), 1–48.

Lazear, E. P. (1999): "Culture and Language," *Journal of Political Economy*, 107(S6), S95–S126.

List, J. A. (2004): "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics*, 119(1), 49–89.

Macrae, C. N., and G. V. Bodenhausen (2000): "Social Cognition: Thinking Categorically about Others," *Annual Review of Psychology*, 51(1), 93–120.

Müller-Itten, M., and A. Öry (2022): "Mentoring and the Dynamics of Affirmative Action," *American Economic Journal: Economic Policy*, forthcoming.

Neckerman, K. M., and J. Kirschenman (1991): "Hiring Strategies, Racial Bias, and Inner-City Workers," *Social Problems*, 38(4), 433–447.

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

Pallais, A., and E. G. Sands (2016): "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy*, 124(6), 1793–1828.

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Eeview*, 62(4), 659–661.

Ritter, J. A., and L. J. Taylor (2011): "Racial Disparity in Unemployment," *Review of Economics and Statistics*, 93(1), 30–42.

Simon, C. J., and J. T. Warner (1992): "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure," *Journal of Labor Economics*, 10(3), 306–330.

Online Appendix

B Supplementary Discussion and Tables

B.1 Sensitivity

In this section we discuss the sensitivity of our results in several dimensions.

Productivity. In Section 4.2 we estimate our baseline productivity regression (11) via weighted-least squares. Tables 3 and 4 shows the resulting supermodularity statistics and confidence intervals from 10,000 bootstraps. As stated in the main text, mean productivity is supermodular, especially for white-Hispanic and black-Hispanic pairs, and productivity variance is log-submodular, especially for black-Hispanic pairs. Thus, the black-Hispanic pairs indicate screening discrimination (97%) while white-Hispanic indicate screening or complementarity (99.1%). Table B1 shows the regression coefficients underlying these results. As discussed in Section 4.2, Table B2 conducts several robustness checks and Figure B2 recalculates Figure 4 for (a) the first six months of workers' tenure, and (b) later months.

Turnover. Section 4.3 considered turnover after 6 months, showing that $\Delta^x_{\Theta\Theta'} < 0$ for all three pairs, with white-Hispanic being significant. The underlying regression is shown in Table B3. Table B4(a) shows that the results are robust to measuring turnover at 3, 9 and 12 months. All the coefficients are negative except black-Hispanic at 6 months; only the white-Hispanic coefficients are significantly negative.

Table B4(b) then includes the worker's average productivity, $\bar{e}_i$, as a control. As expected, this has a negative sign; e.g. with the "6 month" regression, the coefficient is -0.0609, meaning a one standard deviation increase in $\bar{e}_i$ of 0.439 lowers turnover by 2.7% from a base of 36.1%. In the model, a worker is fired when their realized productivity falls below a threshold. Under screening discrimination and complementary production, same-race workers have higher productivity than cross-race workers, so controlling for SPH should increase $\hat{\Delta}^x_{\Theta\Theta}$ towards zero. Under taste-based discrimination, same-race workers have lower productivity than cross-race workers, so controlling for SPH should lower $\hat{\Delta}^x_{\Theta\Theta}$. Table B4(b) shows that $\hat{\Delta}^x_{\Theta\Theta}$ tends to increase for white-Hispanic and black-Hispanic pairs.

Full sample of workers. Our baseline results in Section 4 focus on 48,755 WBH newly hired salespeople for whom we have productivity data. We can also run the hiring regression (9) and 6-month turnover regression (15) on the entire sample of 812,399 WBH newly hired workers that includes cashiers, backroom operators, and service technicians. Table B5 shows our hiring results. As before, managers are significantly more likely to hire same-race workers. Specifically, the gap between same-race and cross-race hiring $\hat{\beta}^r_{\Theta\theta} - \hat{\beta}^r_{\Theta\theta}$ is 2.5pp, 3.2pp and 2.4pp for white, black and Hispanic workers, which is similar to the corresponding gaps for salespeople (3.9pp, 2.9pp and 4.0pp). Table B6 shows turnover is submodular for each pair of races; these results are significant at the 1% level. The size of these coefficients is $(\hat{\Delta}^x_{WB} + \hat{\Delta}^x_{WH} + \hat{\Delta}^x_{BH})/6 = -2.9\%$ compared with -0.8% for salespeople. Table B4(c) shows these results are robust measuring at turnover at 3, 9 and 12 months.

Internal transfers. Our test of discrimination in Table 4 uses the productivity of a worker under their hiring manager. If workers were reassigned randomly (e.g. Glover, Pallais, and Pariente (2017)), then we could use productivity under non-hiring managers to test for complementarity. Unfortunately, this is not the case. First, managers employ about 4.7pp more same-race candidates than one would expect from random matching (albeit with large standard errors). Second, the race of the hiring manager and the new manager are correlated (after controlling for store and department fixed effects). Our random assignment assumption (see Section 2.5) may be satisfied after controlling for the race of a worker's hiring manager (and possibly their productivity under the hiring manager). One could then apply our discrimination tests to internal transfers by comparing the supermodularity of hiring and productivity conditional on the hiring manager's race. Unfortunately we do not have enough transfers to tightly estimate such parameters. In addition, transfers might depend on workers' preferences, which complicates the theory.

B.2 Alternative theories

This paper has considered three theories (taste-based discrimination, screening discrimination, complementary production). Here we briefly discuss other factors that might be at play.

Customer bias. Our main theories suppose that worker productivity is determined by the hiring decisions of their managers (and directly by the manager, in the case of complementarity). If manager race is correlated with the race of customers, biased customer preferences may look like complementary production between the manager and worker. To be specific, if there are "white towns" with predominantly white customers and white managers, and "Hispanic towns" with predominantly Hispanic customers and Hispanic managers, then same-race preferences among customers will induce $\hat{\Delta}^y_{WH} > 0$ even without complementarity between managers and workers.[37] To address this, we re-run our productivity regression (11) and include the fraction of the population of the same race as the worker, which serves as a proxy for the race of customers. Column (2) of Table B8 shows the "share same-race pop" coefficient is positive (1.5pp) but not significant. Including this variable

[37] We allow for store fixed effects, which accounts for the fact that all workers may be more productive in some stores than others. However, in the white/Hispanic town example, the stores might have equal average productivity.

has little impact on the supermodularity statistics, $\hat{\Delta}^y_{\Theta\theta}$ or $\hat{\Delta}^v_{\Theta\theta}$, suggesting that customer bias is not driving the results in Tables 3-4.

Peer effects. Analogous to customer bias, if manager race is correlated with the race of other team members, complementarity between peers may look like complementarity between the manager and worker. To be specific, if white managers are in charge of predominantly white teams and Hispanic managers are in charge of predominantly Hispanic teams, then complementarity between peers will induce $\hat{\Delta}^y_{WH} > 0$ even without complementarity between managers and workers. To address this, we re-run our productivity regression (11) and include the fraction of peers of the same race as the worker. Column (3) of Table B8 shows the "share same-race worker peers" coefficient is positive (1.8pp) and significant at the 10% level. However, including this variable does not qualitatively change the supermodularity statistics ($\hat{\Delta}^y_{\Theta\theta}, \hat{\Delta}^v_{\Theta\theta}$), suggesting that peer effects are not driving the results in Tables 3-4. With this said, the interpretation of this regression is unclear since the race of a worker's peers is endogenous and the positive "worker peers" coefficient may pick up managerial heterogeneity. For example, if a manager has strong same-race signals, the average worker will have more same-race peers and the workers will have high productivity.[38]

Favoritism. Our theory allows for complementarity, whereby a worker is more productive with a same-race manager. This is different from favoritism, whereby a manager gives preferential treatment to same race workers. The latter is intrinsically zero-sum, e.g. assigning same-race workers to the best shifts. Our measure of productivity controls for the worker's shift via their target sales, eliminating the most obvious source of favoritism. In order to test for residual favoritism, we note that if the manager gives the best roles to same-race workers then any given worker's productivity should improve if there are fewer peers who share the same race as the manager. So inspired, we re-run our productivity regression (11) and include the fraction of peers of the same race as the manager. Column (4) of Table B8 shows the "share same-race manager peers" coefficient is positive (2.6pp) and significant at the 5% level; in contrast, the favoritism hypothesis would predict a negative coefficient. This is evidence against favoritism but, as with the peer effects, this result may come from reverse causation since the race of a worker's peers is endogenous and the positive "manager peers" coefficient may pick up managerial heterogeneity. For example, if a manager has strong same-race signals, she will have more same-race peers and her workers will have high productivity.

Referral networks. As discussed in Section 2.5, our model assumes that each manager selects from

[38]One may also wonder if the race of peers are correlated with managers' hiring decisions. Table B7 includes peer race in our baseline hiring regression and shows that both managers' race and peer race is correlated with the race of new hires. Again, the interpretation of this is unclear. Having more peers of one race may bias the manager towards future hires of the same race: For example, Benson and LePage (2022) use the current data set to examine how a manager's hiring decision depends on their past hires. Alternatively, since the mix of peers is endogenous, the positive peer race coefficient may pick up heterogeneity in managerial bias: If a manager is very biased, their teams and new hires will both be more likely to be same-race.

an equal distribution of applicants $(n_\Theta, \mu_\Theta, \sigma^2_\Theta)$ after controlling for store, department and month fixed effects. This excludes the possibility that the manager may obtain referrals through their social networks. If referrals have the same productivity distribution $(\mu_\Theta, \sigma^2_\Theta)$ as other applicants of that race, the larger number of applicants n_Θ would raise the number of same-race hires (Table 2) but would not affect productivity (Tables 3-4). However, referrals may also be positively selected (e.g. Beaman and Magruder (2012), Pallais and Sands (2016)). If the manager receives valuable information from their social contacts, this would lower σ^2_Θ and look like our screening model. Alternatively, if referrals have a higher mean distribution μ_Θ, this would look like our complementary production model.

It is hard to evaluate the importance of referrals since we do not observe applications. At an anecdotal level, one HR representative told us that, in her view, it is rare for managers to know applicants. We can also look for indirect evidence of referrals. Brown, Setren, and Topa (2016) find that "most referrals take place between a provider and a recipient with similar characteristics in terms of age, gender, ethnicity, education, and division and staff level within the corporations." Thus, if more of the same-race workers are referrals, one might expect the same-race workers share other characteristics with the manager. Table B9 looks at the average age and gender differences between the worker and manager by each racial pair; there is no evidence that same-race workers are more similar to their managers.

Worker preferences. The model of taste-discrimination assumes that managers are biased towards same-race workers. An alternative model is that managers are neutral, but workers are biased towards same-race managers, accepting the offer with a higher probability when the manager shares the same race. Such a model also predicts the same managers hire more same-race candidates. The impact on productivity depends on the details of the model. If workers' decision to reject offers are independent of their skill, the resulting productivity is modular, $\Delta^y_{\Theta\theta} = 0$. Alternatively, if workers have a fixed outside option that depends on the race of the manager, then low-productivity workers will be more likely to reject cross-race offers, meaning productivity is submodular. Ultimately, we cannot rule out such forces, but they seem less important with hiring decisions than retention or transfer decisions.

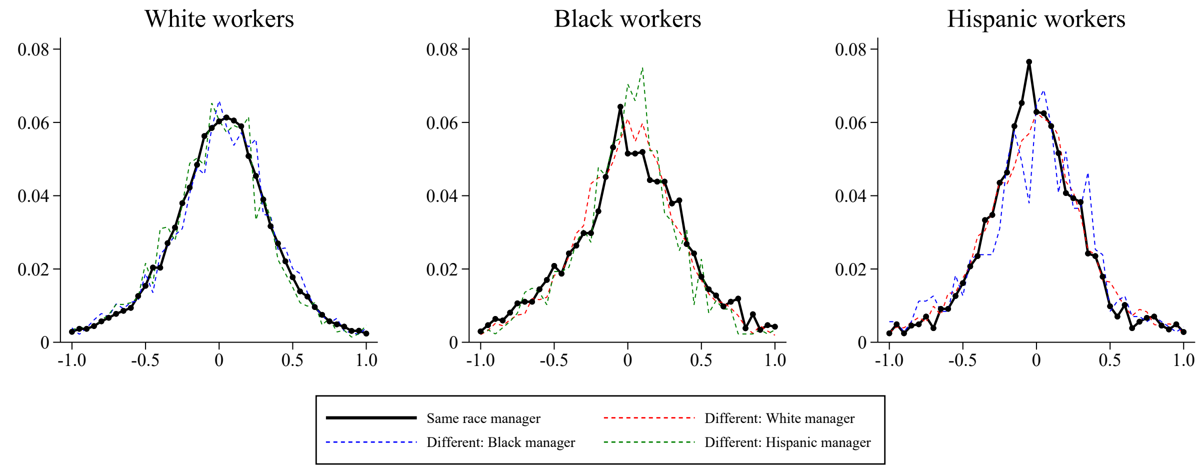


FIGURE B1: **HISTOGRAMS OF WORKER PRODUCTIVITY BY MANAGER AND WORKER RACE**
Notes: Recall that $\bar{e}_i$ is the within-worker mean residual from the productivity regression (11). By construction, $\bar{e}_i$ has zero mean within each race-pair $\Theta\theta$, so $\bar{\varepsilon}_i := \bar{e}_i + \hat{\beta}^y_{\Theta\theta}$ is the productivity of worker i, after controlling for store, department, month and tenure. The figure plots the histogram of $\bar{\varepsilon}_i$ across manager and worker race. For example, the left hand figure shows the distribution of productivity for white workers; the black line is productivity under white managers, the green line is productivity under black managers, and the blue line productivity under Hispanic managers. We plot the frequency distributions for every combination of racial pair at 5%-wide intervals of productivity.

Table B1: **Baseline productivity regression**

Worker-Manage race	Prod. Mean, $\hat{\beta}^y_{\Theta\theta}$	90% Conf. Int.	Prod. Variance, $\hat{\sigma}^2_{\Theta\theta}$	No. workers
White-White	reference		0.114	26,847
White-Black	-0.0019	[-0.0172, 0.0155]	0.120	2,285
White-Hispanic	-0.0260	[-0.0415, -0.0073]	0.111	2,122
Black-White	-0.0251	[-0.0358, -0.0139]	0.143	6,172
Black-Black	-0.0130	[-0.0297, 0.0050]	0.143	2,344
Black-Hispanic	-0.0323	[-0.0572, -0.0045]	0.129	877
Hispanic-White	-0.0163	[-0.0283, -0.0050]	0.119	4,557
Hispanic-Black	-0.0383	[-0.0695, -0.0076]	0.172	710
Hispanic-Hispanic	-0.0146	[-0.0281, 0.0013]	0.105	2,841
Productivity Mean:				
White-Black ($\hat{\Delta}^y_{WB}$)	0.0140	[-0.0117, 0.0383]		37,648
White-Hispanic ($\hat{\Delta}^y_{WH}$)	0.0277	[0.0049, 0.0510]		36,367
Black-Hispanic ($\hat{\Delta}^y_{BH}$)	0.0429	[-0.0011, 0.0880]		6,772
Productivity Variance:				
White-Black ($\hat{\Delta}^v_{WB}$)	0.953	[0.799, 1.131]		37,648
White-Hispanic ($\hat{\Delta}^v_{WH}$)	0.914	[0.759, 1.110]		36,367
Black-Hispanic ($\hat{\Delta}^v_{BH}$)	0.679	[0.500, 0.956]		6,772
Location FEs	Yes			
Month FEs	Yes			
Department FEs	Yes			
Observations	336,046			48,755

Notes: This table shows the outcome of the baseline productivity regression (11) that underlies Tables 3-4, estimated via WLS, inversely weighted to tenure. The top section shows the regression coefficients $\hat{\beta}^y_{\Theta\theta}$ by race-pair and the corresponding cross-agent variance $\hat{\sigma}^2_{\Theta\theta}$ given by (13). The second section shows the mean supermodularity coefficient $\hat{\Delta}^y_{\Theta\Theta'}$ as defined by (12), while the third section shows the variance supermodularity coefficient $\hat{\Delta}^v_{\Theta\Theta'}$ as defined by (14). The confidence intervals come from 10,000 bootstraps.

Table B2: Discrimination tests: Sensitivity

Statistic	WLS-T (1)	WLS-NT (2)	WLS-3T (3)	WLS-6T (4)	OLS-T (5)	OLS-NT (6)	OLS-3T (7)	OLS-6T (8)
Productivity Mean:								
White-Black ($\hat{\Delta}^y_{WB}$)	0.0140	0.0143	0.0181	0.0126	-0.0025	-0.0025	0.0039	-0.0012
	[-0.0117, 0.0383]	[-0.0126, 0.0382]	[-0.0033, 0.0402]	[-0.0072, 0.0326]	[-0.0274, 0.0237]	[-0.0264, 0.0231]	[-0.0208, 0.0300]	[-0.0242, 0.0217]
White-Hispanic ($\hat{\Delta}^y_{WH}$)	0.0277	0.0301	0.0253	0.0212	-0.0012	0.0003	0.0125	0.0105
	[0.0049, 0.0510]	[0.0076, 0.0529]	[0.0049, 0.0450]	[0.0024, 0.0393]	[-0.0240, 0.0236]	[-0.0227, 0.0254]	[-0.0109, 0.0360]	[-0.0111, 0.0313]
Black-Hispanic ($\hat{\Delta}^y_{BH}$)	0.0429	0.0454	0.0283	0.0267	0.0163	0.0179	0.0278	0.0264
	[-0.0011, 0.0880]	[-0.0001, 0.0863]	[-0.0083, 0.0626]	[-0.0103, 0.0641]	[-0.0244, 0.0557]	[-0.0217, 0.0598]	[-0.0151, 0.0720]	[-0.0159, 0.0688]
Productivity Variance:								
White-Black ($\hat{\Delta}^v_{WB}$)	0.953	0.950	0.921	1.005	0.929	0.927	0.915	0.996
	[0.799, 1.131]	[0.797, 1.148]	[0.772, 1.102]	[0.833, 1.185]	[0.780, 1.099]	[0.775, 1.099]	[0.761, 1.095]	[0.846, 1.189]
White-Hispanic ($\hat{\Delta}^v_{WH}$)	0.914	0.915	0.931	0.999	0.911	0.909	0.933	1.004
	[0.759, 1.110]	[0.765, 1.131]	[0.772, 1.138]	[0.822, 1.199]	[0.759, 1.088]	[0.740, 1.089]	[0.765, 1.118]	[0.827, 1.223]
Black-Hispanic ($\hat{\Delta}^v_{BH}$)	0.679	0.680	0.840	0.798	0.664	0.665	0.843	0.800
	[0.500, 0.956]	[0.503, 0.941]	[0.620, 1.195]	[0.592, 1.103]	[0.493, 0.933]	[0.490, 0.939]	[0.616, 1.209]	[0.574, 1.111]
White-Black joint test (%):								
$\hat{\Delta}^y > 0$, $\hat{\Delta}^v < 1$ (screen)	57.9	58.7	73.5	45.1	30.9	34.4	45.6	24.8
$\hat{\Delta}^y > 0$, $\hat{\Delta}^v > 1$ (comple)	23.6	23.1	18.3	41.1	10.3	8.7	13.2	21.8
$\hat{\Delta}^y < 0$, $\hat{\Delta}^v < 1$ (s & t)	10.7	9.6	5.3	4.5	45.8	43.4	31.6	24.0
$\hat{\Delta}^y < 0$, $\hat{\Delta}^v > 1$ (taste)	7.8	8.6	2.9	9.3	13.0	13.5	9.6	29.4
White-Hispanic joint test (%):								
$\hat{\Delta}^y > 0$, $\hat{\Delta}^v < 1$ (screen)	75.6	78.3	71.2	50.7	36.0	40.3	57.7	36.1
$\hat{\Delta}^y > 0$, $\hat{\Delta}^v > 1$ (comple)	22.1	20.7	27.6	46.0	12.3	9.7	24.1	42.4
$\hat{\Delta}^y < 0$, $\hat{\Delta}^v < 1$ (s & t)	1.4	0.8	0.7	1.7	41.6	40.4	12.8	11.0
$\hat{\Delta}^y < 0$, $\hat{\Delta}^v > 1$ (taste)	0.9	0.2	0.5	1.6	10.1	9.6	5.4	10.5
Black-Hispanic joint test (%):								
$\hat{\Delta}^y > 0$, $\hat{\Delta}^v < 1$ (screen)	91.8	92.4	73.5	78.5	72.7	77.1	67.0	75.8
$\hat{\Delta}^y > 0$, $\hat{\Delta}^v > 1$ (comple)	2.6	2.5	17.4	9.8	1.6	1.6	18.3	10.2
$\hat{\Delta}^y < 0$, $\hat{\Delta}^v < 1$ (s & t)	5.2	4.6	6.2	9.1	25.1	20.6	9.9	10.5
$\hat{\Delta}^y < 0$, $\hat{\Delta}^v > 1$ (taste)	0.4	0.5	2.9	2.6	0.6	0.7	4.8	3.5

Notes: This table reports the supermodularity statistics (as in Table 3) and discrimination tests (as in Table 4) for seven model variants. Column (1) is the baseline specification, using weighted-least squares and a tenure control. Column (2) dispenses with the tenure control while Columns (3) and (4) use only the first three and six months of each workers tenure, respectively. Columns (5)-(8) show the corresponding results for ordinary-least-squares. Below each coefficient is the 90% confidence interval computed from 1000 bootstraps (except for Column (1), which uses 10,000 bootstraps).

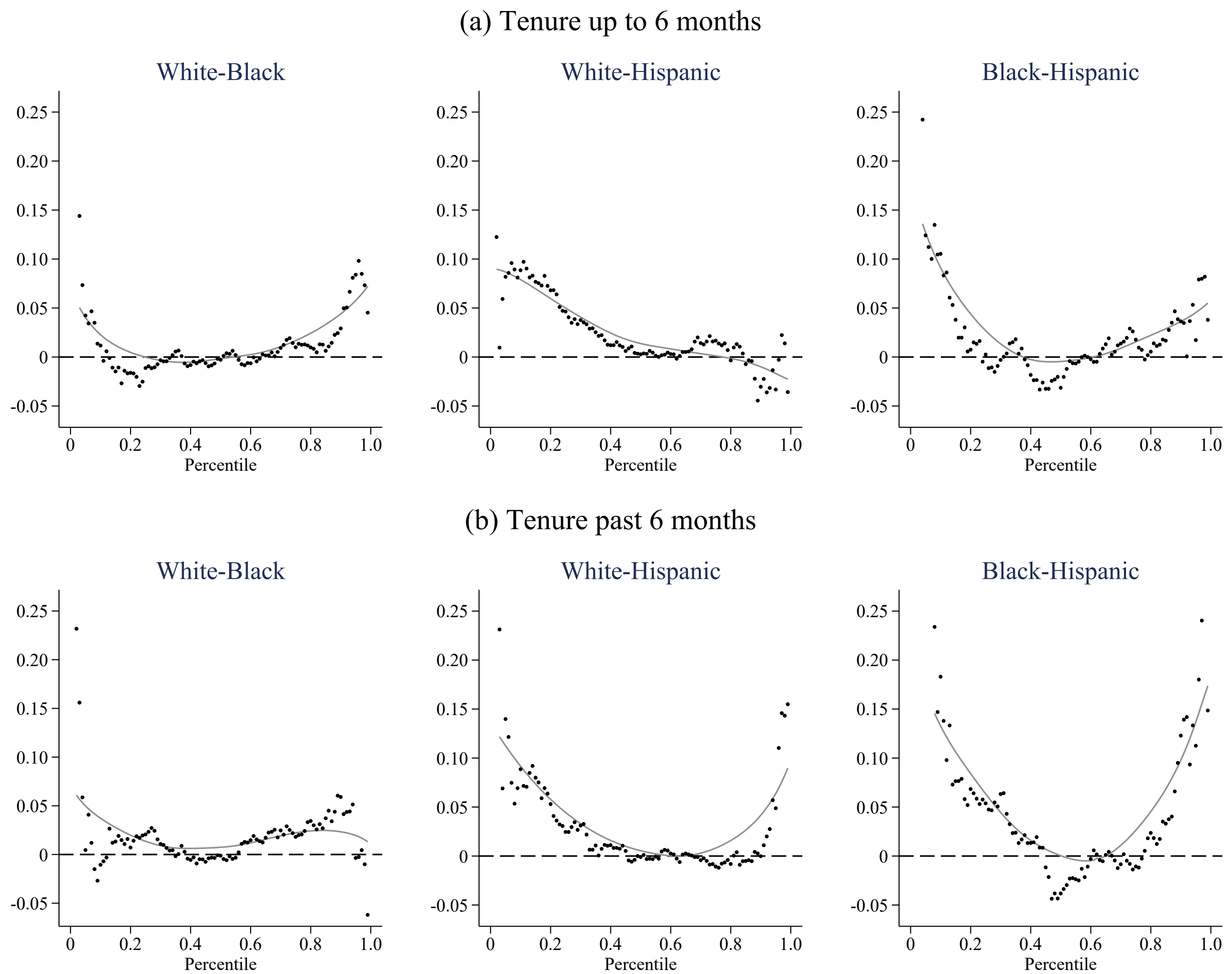


FIGURE B2: **SUPERMODULARITY OF PRODUCTIVITY BY QUANTILE**

Notes: This figure recreates Figure 4 with data above/below six months. Specifically, we run the baseline productivity regression (11) and take the residuals e_{it}. We then truncate the data, calculate i's average residual $\bar{e}_i$ and calculate $\hat{\Delta}^{y^q}_{\Theta\Theta'}$ as defined in (19). This figure shows across 98 one-percentile bins, after we have thrown away the top and bottom bins. The Lowess curve has bandwidth 0.8.

Table B3: **Baseline turnover regression**

6 month turnover	Point estimate	90% confidence interval	Number of workers
White-White ($\hat{\beta}^x_{Ww}$)	reference		21,613
White-Black ($\hat{\beta}^x_{Wb}$)	0.0033	[-0.0155, 0.0221]	1,885
White-Hispanic ($\hat{\beta}^x_{Wh}$)	-0.0128	[-0.0326, 0.0069]	1,694
Black-White ($\hat{\beta}^x_{Bw}$)	0.0425	[0.0302, 0.0548]	5,007
Black-Black ($\hat{\beta}^x_{Bb}$)	0.0304	[0.0116, 0.0491]	1,930
Black-Hispanic ($\hat{\beta}^x_{Bh}$)	-0.0157	[-0.0455, 0.0141]	717
Hispanic-White ($\hat{\beta}^x_{Hw}$)	-0.0073	[-0.0213, 0.0067]	3,699
Hispanic-Black ($\hat{\beta}^x_{Hb}$)	-0.0104	[-0.0436, 0.0228]	572
Hispanic-Hispanic ($\hat{\beta}^x_{Hh}$)	-0.0555	[-0.0733, -0.0378]	2,222
Location FEs	Yes		
Month FEs	Yes		
Department FEs	Yes		
Observations	39,339		
Mean turnover	0.361		
Pairwise relations			
White-Black ($\hat{\Delta}^x_{WB}$)	-0.0154	[-0.0436, 0.0127]	30,435
White-Hispanic ($\hat{\Delta}^x_{WH}$)	-0.0354	[-0.0698, -0.0011]	29,228
Black-Hispanic ($\hat{\Delta}^x_{BH}$)	0.0009	[-0.0589, 0.0607]	5,441

Notes: This table shows the baseline turnover regression (15) that underlies the turnover results in Table 3. The supermodularity coefficient $\hat{\Delta}^x_{WB}$ is defined by (16).

Table B4: **Turnover supermodularity statistics: Sensitivity**

	White-Black ($\hat{\Delta}^x_{WB}$)		White-Hispanic ($\hat{\Delta}^x_{WH}$)		Black-Hispanic ($\hat{\Delta}^x_{BH}$)	
	Point est.	90% conf. int.	Point est.	90% conf. int.	Point est.	90% Conf. Int.
(a) Salespeople						
3 months	-0.0144	[-0.0391, 0.0103]	-0.0130	[-0.0434, 0.0174]	-0.0316	[-0.0841, 0.0209]
6 months	-0.0154	[-0.0436, 0.0127]	-0.0354	[-0.0698, -0.0011]	0.0009	[-0.0589, 0.0607]
9 months	-0.0221	[-0.0520, 0.0077]	-0.0444	[-0.0806, -0.0082]	-0.0293	[-0.0924, 0.0338]
12 months	-0.0224	[-0.0529, 0.0082]	-0.0442	[-0.0813, -0.0071]	-0.0417	[-0.1063, 0.0229]
(b) Salespeople, controlling for productivity						
3 months	-0.0270	[-0.0533, -0.0006]	-0.0004	[-0.0338, 0.0330]	-0.0315	[-0.0874, 0.0244]
6 months	-0.0206	[-0.0507, 0.0094]	-0.0053	[-0.0431, 0.0324]	0.0176	[-0.0460, 0.0813]
9 months	-0.0341	[-0.0660, -0.0021]	-0.0099	[-0.0499, 0.0301]	-0.0136	[-0.0811, 0.0540]
12 months	-0.0284	[-0.0612, 0.0045]	-0.0087	[-0.0498, 0.0324]	-0.0149	[-0.0844, 0.0546]
(c) All workers						
3 months	-0.0289	[-0.0353, -0.0225]	-0.0566	[-0.0645, -0.0487]	-0.0491	[-0.0621, -0.0361]
6 months	-0.0382	[-0.0458, -0.0307]	-0.0730	[-0.0823, -0.0636]	-0.0641	[-0.0795, -0.0487]
9 months	-0.0379	[-0.0454, -0.0304]	-0.0887	[-0.0979, -0.0795]	-0.0789	[-0.0941, -0.0637]
12 months	-0.0314	[-0.0385, -0.0243]	-0.0945	[-0.1033, -0.0857]	-0.0800	[-0.0944, -0.0656]

Notes: This table shows how the supermodularity coefficient for the turnover regression (15) depends on our assumptions. Part (a) shows the regression for salespeople where $TURN_i$ is classified as leaving after 3, 6, 9 and 12 months. The "6 month" row corresponds to the baseline regression in Table B3. Part (b) shows the same regressions for salespeople, where we add the person's average productivity $\bar{e}_i$ as a control variable, as defined in Section 3.1. The coefficient on $\bar{e}_i$ is stable across the regressions; for example, in the "6 month" model the coefficient on $\bar{e}_i$ is -0.0609 with a standard error of 0.0060. Part (c) shows the same regression as part (a), for all workers rather than just salespeople.

Table B5: **Hiring tests for all workers**

	White worker	Black worker	Hispanic worker
White manager	57.5 [57.4, 57.6]	25.2 [25.1, 25.3]	17.3 [17.2, 17.4]
Black manager	54.5 [54.2, 54.9]	28.0 [27.7, 28.3]	17.5 [17.2, 17.7]
Hispanic manager	55.6 [55.3, 56.0]	24.5 [24.2, 24.8]	19.8 [19.6, 20.1]

Notes: This table shows the hiring regression (9) for 812,399 newly hired WBH workers hired by WBH managers, controlling for location, department, and month. Below each coefficient is the 90% confidence interval.

Table B6: **Turnover regression for all workers**

6 month turnover	Point estimate	90% confidence interval	Number of workers
White-White ($\hat{\beta}^x_{Ww}$)	reference		331,665
White-Black ($\hat{\beta}^x_{Wb}$)	0.0548	[0.0521, 0.0575]	119,984
White-Hispanic ($\hat{\beta}^x_{Wh}$)	-0.0045	[-0.0079, -0.0011]	66,017
Black-White ($\hat{\beta}^x_{Bw}$)	0.0008	[-0.0047, 0.0064]	22,250
Black-Black ($\hat{\beta}^x_{Bb}$)	0.0174	[0.0125, 0.0222]	29,996
Black-Hispanic ($\hat{\beta}^x_{Bh}$)	-0.0378	[-0.0468, -0.0292]	8,490
Hispanic-White ($\hat{\beta}^x_{Hw}$)	-0.0100	[-0.0159, -0.0041]	19,489
Hispanic-Black ($\hat{\beta}^x_{Hb}$)	0.0320	[0.0247, 0.0394]	12,364
Hispanic-Hispanic ($\hat{\beta}^x_{Hh}$)	-0.0874	[-0.0918, -0.0831]	37,693
Location FEs	Yes		
Month FEs	Yes		
Department FEs	Yes		
Observations	648,036		
Mean turnover	0.489		
Pairwise relations			
White-Black ($\hat{\Delta}^x_{WB}$)	-0.0382	[-0.0458 , -0.0307]	503,895
White-Hispanic ($\hat{\Delta}^x_{WH}$)	-0.0730	[-0.0808 , -0.0651]	454,864
Black-Hispanic, ($\hat{\Delta}^x_{BH}$)	-0.0641	[-0.0770 , -0.0512]	88,543

Notes: This table shows the turnover regression (15) for all newly hired WBH workers hired by WBH managers.

Table B7: **Hiring tests with peer race**

	(1) White worker	(2) Black worker	(3) Hispanic worker
White manager	ref	-2.5 [-3.6, -1.4]	-4.0. [-5.0, -3.0]
Black manager	-2.8 [-4.1, -1.5]	ref.	-3.8 [-5.1, -2.5]
Hispanic manager	-4.3 [-5.6, -2.9]	-2.3 [-3.7, -0.9]	ref.
White peers	4.7 [-1.3, 10.6]	-2.7 [-7.7, 2.4]	-2.0 [-6.6, 2.6]
Black peers	-1.7 [-8.4, 4.9]	5.4 [-0.3, 11.0]	-3.6 [-8.7, 1.4]
Hispanic peers	1.5 [-5.3, 8.4].	-1.5 [-7.3, 4.3]	0.0 [-5.3, 5.2]
Constant	62.5 [56.9, 68.0]	22.5 [17.8, 27.3]	21.5 [17.2, 25.9]

Notes: This table expands the hiring regression (9) to include the share of white/black/Hispanic peers. As in the baseline hiring regression, we allow for store, department and month fixed effects. There are 43,164 salespeople; this is fewer than in Table 2 since we only observe team composition for people hired during the sample period. Below each coefficient is the 90% confidence interval.

Table B8: **Discrimination tests: Race of customers and peers**

Statistic	Baseline (1)	Customers (2)	Worker Peers (3)	Manager Peers (4)
Productivity Mean:				
White-Black ($\hat{\Delta}^y_{WB}$)	0.0140	0.0112	0.0057	0.0028
	[−0.0117, 0.0383]	[-0.0130, 0.0352]	[-0.0186, 0.0319]	[-0.0220, 0.0284]
White-Hispanic ($\hat{\Delta}^y_{WH}$)	0.0277	0.0232	0.0179	0.0093
	[0.0049, 0.0510]	[-0.0009, 0.0460]	[-0.0063, 0.0435]	[-0.0140, 0.0349]
Black-Hispanic ($\hat{\Delta}^y_{BH}$)	0.0429	0.0396	0.0338	0.0304
	[−0.0011, 0.0880]	[-0.0034, 0.0863]	[-0.0096, 0.0785]	[-0.0122, 0.0727]
Productivity Variance:				
White-Black ($\hat{\Delta}^v_{WB}$)	0.953	0.951	0.948	0.923
	[0.799, 1.131]	[0.806, 1.120]	[0.797, 1.120]	[0.771, 1.097]
White-Hispanic ($\hat{\Delta}^v_{WH}$)	0.914	0.804	0.910	0.933
	[0.759, 1.110]	[0.662, 0.978]	[0.749, 1.098]	[0.770, 1.122]
Black-Hispanic ($\hat{\Delta}^v_{BH}$)	0.679	0.594	0.682	0.709
	[0.500, 0.956]	[0.435, 0.830]	[0.496, 0.954]	[0.516, 1.003]
White-Black joint test (%):				
$\Delta^y > 0$, $\Delta^v < 1$ (screen)	57.9	56.0	49.7	49.0
$\Delta^y > 0$, $\Delta^v > 1$ (comple)	23.6	21.0	17.2	10.2
$\Delta^y < 0$, $\Delta^v < 1$ (s & t)	10.7	14.3	19.3	26.8
$\Delta^y < 0$, $\Delta^v > 1$ (taste)	7.8	8.7	13.8	14.0
White-Hispanic joint test (%):				
$\Delta^y > 0$, $\Delta^v < 1$ (screen)	75.6	91.4	70.7	55.7
$\Delta^y > 0$, $\Delta^v > 1$ (comple)	22.1	3.0	17.5	20.0
$\Delta^y < 0$, $\Delta^v < 1$ (s & t)	1.4	5.5	9.7	16.7
$\Delta^y < 0$, $\Delta^v > 1$ (taste)	0.9	0.1	2.1	7.6
Black-Hispanic joint test (%):				
$\Delta^y > 0$, $\Delta^v < 1$ (screen)	91.8	92.7	88.3	83.9
$\Delta^y > 0$, $\Delta^v > 1$ (comple)	2.6	0.6	2.7	4.1
$\Delta^y < 0$, $\Delta^v < 1$ (s & t)	5.2	6.5	8.7	10.9
$\Delta^y < 0$, $\Delta^v > 1$ (taste)	0.4	0.2	0.3	1.1
Share Same-Race Population		0.0145		
		[-0.0103, 0.0425]		
Share Same-Race Worker Peers			0.0184	
			[0.0026, 0.0334]	
Share Same-Race Manager Peers				0.0255
				[0.0111 0.0397]

Notes: This table reports the supermodularity statistics (as in Table 3) and discrimination tests (as in Table 4) for three model variants. Column (1) repeats the baseline results. Column (2) proxies for the race of customers by controlling for the share of the local population who are the same race as the worker that year. We use MSA population, or non-MSA state population if the location isn't in an MSA, from the annual American Community Survey. Column (3) controls for the share of the worker's peers who share the same race as the worker that month (taking into account non-WBH races). Column (4) controls for the share of worker's peers who share the same race as the manager that month (taking into account non-WBH races). We present 90% confidence intervals calculated via 1,000 bootstraps.

TABLE B9: **REFERRALS TEST: AGE AND GENDER DIFFERENCES BY RACIAL PAIR**

	Age difference in years			Same gender indicator		
	White worker	Black worker	Hispanic worker	White worker	Black worker	Hispanic worker
White manager	17.20	17.54	18.42	0.383	0.381	0.386
	(11.30)	(11.50)	(11.51)	(0.486)	(0.486)	(0.487)
	[17.09, 17.32]	[17.29, 17.78]	[18.13, 18.70]	[0.378, 0.388]	[0.371, 0.391]	[0.373, 0.398]
Black manager	16.78	17.05	17.33	0.382	0.411	0.392
	(10.77)	(11.17)	(10.16)	(0.486)	(0.492)	(0.489)
	[16.40, 17.15]	[16.66, 17.43]	[16.69, 17.97]	[0.365, 0.399]	[0.394, 0.428]	[0.361, 0.423]
Hispanic manager	14.82	13.75	14.66	0.386	0.411	0.426
	(10.24)	(9.44)	(9.15)	(0.487)	(0.492)	(0.495)
	[14.45, 15.19]	[13.22, 14.29]	[14.38, 14.95]	[0.369, 0.404]	[0.383, 0.439]	[0.411, 0.442]
	Cross-race mean		17.12	Cross-race mean		0.385
	Same-race mean		16.96	Same-race mean		0.389
	p-value, same means		0.160	p-value, same means		0.454

Notes: The table reports differences in ages and gender by racial pair, for salespeople in the main sample, given the race of the worker and manager. The **left panel** reports the means and standard deviation for the absolute value of difference in ages in years, along with the 90% confidence interval. For instance, the top left cell restricts the sample to white workers hired by white managers, then reports the mean absolute value of their age differences. The "cross-race mean" reports the average of the "cross-race" cells, weighted by sample size, while the "same-race" mean reports the average of the same-race cells. The p-values correspond to two-sample t-test for equal means, where the dependent variable (the absolute value of the difference in ages) depends on whether the manager is same-race or cross-race. The **right panel** reports the mean and standard deviation of an indicator that equals one if the manager and worker have different genders. A majority of employees are men, so the probability that sexes differ is less than 0.5. Data in this table include 47,094 newly hired sales workers included in the main sample, and exclude approximately 3.5% of workers for whom manager gender or age are missing.

C Proofs

C.1 Maximizing expected sales

In this appendix we analyze the model variant in which managers maximize expected sales, as discussed in Section 2.5. Let Y be a worker's sales, and assume a manager wishes to hire workers with $Y \geq Y^*$. As in the main text, let $\hat{y} = E[y|\tilde{y}]$ be expected log-sales, with mean μ, estimator variance $\eta^2 = \sigma_0^4/(\sigma_0^2 + \sigma_\epsilon^2)$, and estimation error $y - \hat{y}$ with residual variance $\gamma^2 := \sigma_0^2\sigma_\epsilon^2/(\sigma_0^2 + \sigma_\epsilon^2)$.

The manager's expected utility from a worker given estimate $\hat{y}$ is thus

$$E[Y|\hat{y}] = E[e^y|\hat{y}] = e^{\hat{y}}E[e^{y-\hat{y}}|\hat{y}] = e^{\hat{y}+\gamma^2/2}.$$

The manager thus wishes to hire a workers with estimate $\hat{y} \geq y^* := \log Y^* - \gamma^2/2$. Intuitively, managers lower their standard, $y^* < \log Y^*$, since they are risk-loving with respect to uncertainty in log-productivity.

Taste-based discrimination. In this model, biased managers gain extra utility from same-race candidates. Suppose they gain utility Y/D from a cross-race candidate, and Y from a same-race candidate, where $D > 1$ is the same-race bias. Taking logs, the cross-race cutoff $y_c^* = \log Y^* + d - \gamma^2/2$ thus exceeds the same-race cutoff $y_s^* = \log Y^* - \gamma^2/2$ by the bias $d = \log D$. The analysis of hiring is thus identical to that in Section 2.1, and thus the proofs of Propositions 1(a)-(c) carry over unchanged. When it comes to turnover, suppose that the manager fires the worker if their realized productivity y drops more than $\tilde{\tau}$ below the cutoff y_i^*. Thus the turnover rate for $i \in \{c, s\}$ is

$$x_i := \Pr(y \leq y_i^* - \tilde{\tau}|\hat{y} \geq y_i^*) = \Pr(y \leq y_i^* - \tau|\hat{y} \geq y_i^*).$$

where $\tau := \tilde{\tau} - \gamma^2/2$. This is the same as in Section 2.1, and Proposition 1(d) immediately applies.

Complementary production. As with taste-based discrimination, the analysis of the baseline case (where managers maximize expected log-sales) is essentially unchanged. Since the residual variance is fixed at γ^2, the cross-race cutoff y_c^* exceeds the same-race cutoff y_s^* by k, and the firing threshold is τ below the hiring threshold. Proposition 3 immediately generalizes.

Screening discrimination. In this model, the signal variance σ_ϵ^2 is lower for same-race applicants than for cross-race applicants, $\sigma_s^2 \leq \sigma_c^2$. This implies that cross-race hires have more residual variance than same-race hires, $\gamma_s^2 \leq \gamma_c^2$. Thus, managers set a lower threshold for cross-race hires, $y_c^* = \log Y^* - \gamma_c^2/2 < \log Y^* - \gamma_s^2/2 = y_s^*$. Intuitively, managers are risk-loving with respect to the log-productivity, and thus likes cross-race candidates because of their high residual variance. We first verify that managers still hire more same-race workers.

Proposition 2(a)′. *Assume* $\log Y^* > \mu + \sigma_0^2$. *In the screening model, the hiring probability is higher for same-race applicants than cross-race applicants,* $p_s > p_c$.

Proof. For part (a), let Φ and ϕ be the cdf and pdf of the standard normal distribution. Write the precision of the prior as $h_0 := 1/\sigma_0^2$ and the precision of the raw signal as $h_\epsilon = 1/\sigma_\epsilon^2$. The estimator variance and residual variance are then $\eta^2 = h_\epsilon/((h_0+h_\epsilon)h_0)$ and $\gamma^2 = 1/(h_0+h_\epsilon)$. Now, the acceptance rate can be written as a function of the signal precision

$$p(h_\epsilon) := \Pr\left(\hat{y} \geq \log Y^* - \frac{\gamma^2}{2}\right) = 1 - \Phi\left(\frac{\log Y^* - \mu - \frac{\gamma^2}{2}}{\eta}\right) =$$

$$= 1 - \Phi\left(\sqrt{\frac{(h_0+h_\epsilon)h_0}{h_\epsilon}}(\log Y^* - \mu) - \frac{1}{2}\sqrt{\frac{h_0}{(h_0+h_\epsilon)h_\epsilon}}\right)$$

Simple computations show that the derivative equals

$$p'(h_\epsilon) = -\phi(\cdot) \times \sqrt{h_0}\left(-\frac{h_0}{2h_\epsilon}\sqrt{\frac{1}{h_\epsilon(h_0+h_\epsilon)}}(\log Y^* - \mu) + \frac{1}{4}\left(\frac{1}{h_\epsilon} + \frac{1}{h_0+h_\epsilon}\right)\sqrt{\frac{1}{(h_0+h_\epsilon)h_\epsilon}}\right)$$

$$= \phi(\cdot) \times \frac{1}{2}\sqrt{\frac{h_0}{(h_0+h_\epsilon)h_\epsilon}}\frac{h_0}{h_\epsilon}\left((\log Y^* - \mu) - \frac{1}{2}\frac{1}{h_0}\left(1 + \frac{h_\epsilon}{h_0+h_\epsilon}\right)\right)$$

Given that $h_\epsilon/(h_0+h_\epsilon) \leq 1$, this last term is positive if $\log Y^* \geq \mu + \sigma_0^2$. Since same-race candidates have greater signal precision h_ϵ, they have a higher acceptance rate, and so managers hire relatively more same-race workers. □

Turning to productivity, the proof of Proposition 2(b),(c) shows that productivity of cross-race hires has lower mean and greater variance when using the same bar, $y_c^* = y_s^*$. When risk-neutral managers lower the bar for cross-race recruits, this further lowers their expected productivity and raises productivity variance. Finally, when we consider turnover, the proof of Proposition 2(d) shows that cross-race hires have higher turnover when using the same bar, $y_c^* = y_s^*$. When we lower the bar for cross-race recruits, this further raises their turnover.

C.2 Proof of Proposition 6

Let p_s and p_c be the probability same-race and cross-race candidates are hired. Let $M_t \in [0,1]$ and $N_t \in [0,1]$ be the number of A managers and workers at a firm, and consider the dynamics (N_t, M_t) in (N, M) space (see Figure C1).

First, consider the dynamics of workers, given the the number of managers. Let ψ be the rate at which workers and managers leave; this could be due to exit or promotion. A race-A manager

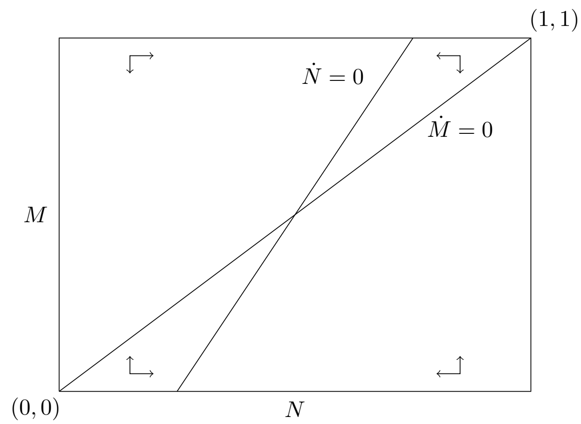


FIGURE C1: **POPULATION DYNAMICS OF MANAGERS AND WORKERS**

hires race-A workers over race-B workers in the ratio $p_s n/p_c(1-n)$. Hence the number of of race-A workers evolves according to

$$\dot{N} = \psi \left[(1-M) \left(\frac{p_c n}{p_c n + p_s(1-n)} \right) + M \left(\frac{p_s n}{p_s n + p_c(1-n)} \right) - N \right]. \tag{31}$$

Thus, in (N, M) space, $\dot{N} = 0$ is a straight line from $(\frac{p_c n}{p_c n + p_s(1-n)}, 0)$ to $(\frac{p_s n}{p_s n + p_c(1-n)}, 1)$, so with slope greater than one.

Next, consider the dynamics of managers, given the number of workers. Managers retire at rate ψ and replacements are chosen randomly from the population of workers, so $\dot{M} = \psi(N - M)$. Thus, $\dot{M} = 0$ along the diagonal where $N = M$.

Figure C1 illustrates the dynamics of (N, M). It is evident that there is a unique steady state where $\dot{M} = 0$ and $\dot{N} = 0$, and that the system converges to this steady state from any initial condition. When $N = M = n$, then

$$\dot{N} = \psi \left[(1-n) \left(\frac{p_c n}{p_c n + p_s(1-n)} \right) - n \left(\frac{p_c(1-n)}{p_s n + p_c(1-n)} \right) \right] \overset{sgn}{=} (p_s - p_c)(2n-1) > 0$$

if $n > 1/2$. Thus the steady state must satisfy $N^* = M^* > n$.

Finally, we consider how many managers of a particular race are required to obtain equitable hiring. Rearranging (31), if we have $\dot{N} = 0$ when $N = n$, then M must solve

$$M \left(\frac{(p_s^2 - p_c^2) n (1-n)}{(p_s n + p_c(1-n)) (p_c n + p_s(1-n))} \right) = \frac{(p_s - p_c)(1-n)n}{p_c n + p_s(1-n)}.$$

Letting $1 - n = \epsilon \approx 0$, we obtain

$$M \left(\frac{p_s^2 - p_c^2}{p_s p_c} \right) \epsilon \approx \frac{p_s - p_c}{p_c} \epsilon,$$

so that the number of race-B managers is $1 - M \approx p_c/(p_s + p_c)$.