NBER WORKING PAPER SERIES

SYSTEMIC DISCRIMINATION: THEORY AND MEASUREMENT

J. Aislinn Bohren
Peter Hull
Alex Imas

Working Paper 29820
http://www.nber.org/papers/w29820

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2022, Revised February 2023

We thank Jason Abaluck, David Arnold, Marianne Bertrand, Peter Blair, Leonardo Bursztyn, Hanming Fang, Damon Jones, Diag Davenport, Will Dobbie, Ed Glaeser, Sam Hirshman, Larry Katz, Erzo Luttmer, Matthew Knepper, Matthew Notowidigdo, Matt Lowe, Anna Gifty Opoku-Agyeman, Jane Risen, Heather Sarsons, Andrei Shleifer, Fabio Tufano, Damián Vergara, Basit Zafar, and seminar participants and conference participants at various institutions for helpful comments. Joshua Hascher, Christy Kang and Cuimin Ba provided expert research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Systemic Discrimination: Theory and Measurement
J. Aislinn Bohren, Peter Hull, and Alex Imas
NBER Working Paper No. 29820
March 2022, Revised February 2023
JEL No. D63,D83,J16,J71

ABSTRACT

Economics often defines and measures discrimination as disparities stemming from direct effects of group identity. We develop new tools to model and measure systemic discrimination, defined as disparities stemming from differences in non-group characteristics. Systemic discrimination can arise from differences in signaling technologies and opportunities for skill development. We propose a measure based on a decomposition of total discrimination into direct and systemic components. The measure is illustrated in a series of hiring experiments and a novel Iterated Audit experimental paradigm with real hiring managers. Results highlight how direct discrimination in one domain can drive systemic discrimination in other domains.

J. Aislinn Bohren
Department of Economics
The Ronald O. Perelman Center for
Political Science and Economics
University of Pennsylvania
133 South 36th Street
Philadelphia, PA 19104
abohren@gmail.com

Peter Hull
Department of Economics
Box B, Brown University
Providence, RI 02912
and NBER
peter_hull@brown.edu

Alex Imas
Booth School of Business
University of Chicago
5897 S. Woodlawn Avenue
Chicago, IL 60637
and NBER
alex.imas@chicagobooth.edu

1 Introduction

Disparities in treatments and outcomes across protected characteristics, such as race and gender, have been widely documented in many settings. Prominent examples include group-based disparities in labor markets, housing, criminal justice, education, and healthcare.[1] In economics, both theoretical and empirical analyses of such disparities tend to focus on the possibility of *direct discrimination*: differential treatment on the basis of the protected characteristic itself, holding other non-group characteristics fixed. For example, a study of labor market discrimination may examine interview call-back rates as a function of perceived race while holding fixed other characteristics, such as education or employment history. Models of how perceived race and gender affect outcomes through people's preferences and beliefs—such as those with taste-based or statistical discrimination (Becker 1957; Phelps 1972; Bohren, Haggag, Imas, and Pope 2022)—have been the primary theoretical tools for studying the drivers of discrimination in economics. The empirical literature has largely followed suit, developing and applying methods to measure the causal effect of protected characteristics on individual and institutional decision-making.

However, a large body of work across other fields has taken a broader view of discrimination. Sociologists and legal scholars have long examined disparities through a systems-based approach, in which racial and gender disparities are seen as a cumulative outcome of both direct and indirect interactions across different stages and domains (Pincus 1996; Powell 2007; De Plevitz 2007). Work on stratification economics argues that observed disparities are due to the incentives of the dominant group to maintain systems of advantage, where discrimination in one domain perpetuates inequity in others (Darity and Mason 1998; Darity 2005). More recently, computer scientists have considered how disparities in algorithmic treatments can arise indirectly from biased data collection and training systems (Angwin, Larson, Mattu, and Kirchner 2016; Rambachan and Roth 2020). From these perspectives, analyses of direct discrimination that condition on non-group characteristics fail to capture the full scope of inequity: such factors may themselves "bake-in" discrimination through interactions with other individuals, markets, and domains.

To illustrate the limits of solely focusing on direct discrimination, consider a stylized labor market example. A recruiter discriminates against female job candidates by giving them lower wage offers than male candidates with identical qualifications. After workers are hired, a manager makes promotion decisions based on performance and salary histories only. Unless the manager considers and adjusts for the recruiter's bias, seemingly non-discriminatory (even gender-neutral) promotion rules will tend to lead to worse outcomes

[1] Examples from these five settings include (i) Gorman (2005), Darity and Mason (1998), Blau and Kahn (2017); (ii) Charles and Hurst (2002), Rugh and Massey (2010), Bayer, Ferreira, and Ross (2017), Yinger (1995); (iii) Mustard (2001), Rehavi and Starr (2014), Arnold, Dobbie, and Hull (2022); (iv) Welch (1973), Card and Krueger (1992), Farkas (2003); and (v) Nazroo (2003), Chandra and Staiger (2010).

for female workers. That is, even if the manager does not *directly* discriminate against female workers conditional on work histories, female workers will be disadvantaged because they have lower salaries. Such *systemic* discrimination is due to gender-based differences in the distribution of salaries (the non-group characteristic), conditional initial qualifications.[2]

A real-world example of this phenomenon comes from *Griggs v. Duke Power Co. (1970)*, a landmark Supreme Court decision on the interpretation of Title VII of the U.S. Civil Rights Act. Griggs argued that Duke Power's policy of requiring a high school diploma for within-company transfers was discriminatory because it disadvantaged Black employees who were otherwise qualified but lacked a degree, in part due to existing discrimination in secondary education. The Court agreed, noting that the high school degree requirement bore no relevance to an individual's ability to perform different jobs at the firm. Notably, discrimination was found despite the transfer policy being facially race-neutral: white and Black employees with the same educational background had the same ability to transfer jobs. Standard economic measures that condition on educational background would have failed to capture the discrimination faced by white and Black workers with the same qualification (i.e., the ability to perform a specific job). Models of taste-based or statistical discrimination would similarly be inappropriate for describing this indirect form of discrimination.[3]

This paper develops new tools to both model and measure such discrimination. We first develop a simple theoretical framework to distinguish *direct* discrimination—differential treatment on the basis of group identity itself—and *systemic* discrimination. We define the latter form as disparities arising indirectly through race- or gender-based differences in the distribution of non-group characteristics among equally-qualified individuals (e.g., differences in performance and salary histories among equally-productive workers). Both forms contribute to *total* discrimination: treatment disparities among equally-qualified individuals. This framework can be used to study different sources of systemic discrimination conditional on the researcher-selected measure of qualification. In the case of *Griggs*, for example, a researcher can align their analysis with the court's by considering disparities conditional on a workers' productivity at Duke Power. Broader notions of systemic discrimination can be obtained by conditioning on upstream measures of qualification (or not conditioning at all), thereby accounting for any systemic factors affecting the worker's current productivity itself.[4]

[2]While in this example direct and systemic discrimination both disadvantage female workers, in principle these two forces could act in opposite directions. We formally outline such a model in Section 4.4.1.

[3]*Griggs* laid the foundation for the disparate impact standard, which considers policies that lead to group-based disparities in outcomes regardless of whether they are neutral with respect to the protected group. This standard is used in a host of contexts, including employment. We discuss the connections between disparate impact and our measures of discrimination below.

[4]Our framework considers direct and systemic discrimination at both the individual and institutional level, and is microfounded by different behavioral and informational structures. Individual direct discrimination can arise from accurate statistical discrimination or from biases in preferences and beliefs. Institutional direct discrimination is generated through the aggregation of individual direct discrimination. Systemic dis-

Our framework makes clear that the study of discrimination requires the researcher to take a normative stance and highlights the value of making this choice explicit for interpreting results and forming any appropriate policy response.

The framework lets us distinguish between two main sources of systemic discrimination. *Informational* systemic discrimination arises due to differences in the process that generates non-group, decision-relevant signals (e.g., of productivity) for the task at hand. This type of systemic discrimination can take the form of signal inflation, in which some signals are systemically higher for one group over the other, or other properties of the signal generating process such as disparities in informativeness due to screening. *Technological* systemic discrimination, in contrast, arises from differences in the relevant productivity measure itself—for example through differences in opportunities for human capital development. We illustrate these drivers in a series of theoretical applications, which show how direct discrimination can have widespread and long-term consequences through systemic discrimination both dynamically and contemporaneously across markets and domains.

We then propose a new measure of systemic discrimination, based on novel Kitagawa-Oaxaca-Blinder decompositions of total discrimination into direct and systemic components.[5] Direct discrimination measures the causal effect of perceived group membership on an action, holding fixed all observable non-group characteristics. Total discrimination is identified by disparities which hold fixed a particular researcher-chosen qualification metric, incorporating both direct discrimination and systemic discrimination through other non-group characteristics. We discuss the types of methods and data that can identify both forms of discrimination. We then propose a general experimental approach, termed an *Iterated Audit* (IA), which can be used to measure systemic discrimination from the decomposition when the qualification measure is observed. We discuss how additional (quasi-)experimental variation can be used to identify or bound systemic discrimination when the qualification metric is only selectively observed or when it can be reliably predicted by observables.

We illustrate our theoretical and empirical frameworks in three experiments. The first two experiments use a stylized lab setting to show how systemic discrimination can arise from signal inflation or differences in signal informativeness. Participants were randomized into one of three roles: Worker, Recruiter, and Hiring Manager. Workers completed two tasks consisting of questions on different subjects. Recruiters observed Worker performance on one task and the Worker's self-reported gender identity. Recruiters then took an action which, along with the Workers' performance on the other task, determined the Recruiters' payoff

---

crimination can arise from disparities in the interactions of individuals or institutions over time, or across different domains within the same time period.

[5] Kitagawa (1955), Oaxaca (1973), and Blinder (1973) decompositions are typically used to measure direct discrimination as the residual of an unconditional disparity, after accounting for differences in observables. Our decomposition instead measures systemic discrimination as the residual of a measure of total discrimination, after accounting for direct discrimination.

via an incentive-compatible mechanism. Hiring Managers also evaluated Workers and took actions after observing Worker gender and a performance signal. But—critically—Managers' signals were determined endogenously through Recruiter actions, allowing direct discrimination by Recruiters to generate systemic discrimination through gender-based differences in the signaling technology. The first study explored systemic discrimination due to signal inflation: Recruiters made wage offers which were then passed along as a (potentially biased) signal to Managers. The second study instead explored systemic discrimination due to differences in signal informativeness: Recruiters made binary hiring decisions, and Managers only observed objective signals of productivity if the Worker was hired by a Recruiter.

Both of these studies revealed significant direct and systemic discrimination. Recruiters made lower wage offers to female Workers than male Workers with similar performance signals, and were less likely to hire female Workers than equally-qualified male Workers. Since there were no gender differences in actual Worker performance, these disparities represent direct discrimination—either due to Recruiter preferences or inaccurate beliefs and stereotypes (Bordalo, Coffman, Gennaioli, and Shleifer 2019; Bohren, Imas, and Rosenberg 2019). There was substantial total discrimination in the behavior of Hiring Managers; male Workers systematically received higher wage offers than equally qualified female Workers.

Our decomposition shows that systemic discrimination played an outsized role in driving these disparities. In the first study, Managers paid female Workers less than male Workers with the same productivity signals (Recruiter wage offers). But direct discrimination by Recruiters led to large gender-based disparities in these signals. Male Workers were paid higher salaries, and this inflation generated systemic disparities that drove the majority of total discrimination. In the second study, Recruiter discrimination in the availability of objective productivity signals led to systemic discrimination in Manager actions: Managers had less information about female candidates because they were less likely to observe the performance signal of a female Worker than of an equally qualified male, which led to subsequent disparities. Moreover, as predicted by our theory, systemic discrimination had a heterogeneous impact depending on Worker qualifications. Disparities in signal informativeness hurt high-performing women who would benefit more from informative signals, but helped low-performing women. Together, these findings show how standard measures of discrimination that condition on non-group characteristics miss large and heterogeneous sources of inequity.

Our third lab-in-the-field experiment used the IA method for detecting systemic discrimination. We recruited a set of actual Hiring Managers with experience evaluating applicants for entry level jobs. In an incentivized factorial ratings design (Kessler, Low, and Sullivan 2019; Lahey and Oxley 2021; Kübler, Schmid, and Stüber 2018), the hiring managers evaluated resumes by reporting hiring likelihoods for an entry-level job at their company. Unlike a standard correspondence or audit study that presents evaluators with two sets of resumes, differing only on randomized signals of group identity, our IA design featured three sets of resumes. Two of the three sets were generated using results of a previous audit study by Pager (2003) who found Black applicants were significantly less likely to proceed through an entry-level job application process than white applicants with equal qualifications. We generate work experience entries in resumes such that the frequency of entries on resumes with distinctively white and Black names matched the disparities generated by direct discrimination in Pager's study; we term the former (latter) set of resumes white(Black)-endogenous. The third set of resumes had the same distribution of work experience as the white-endogenous resumes but featured distinctively Black names; we label this third set of resumes as Black-exogenous. Importantly, in order to give evaluators an opportunity to account for systemic discrimination, Managers were informed of potential race-based disparities in labor market call-back rates prior to making their decisions.

Comparing evaluations of the two endogenous sets of resumes yields a measure of total discrimination, while comparing white-endogenous and Black-exogenous resumes (the standard comparison in correspondence studies) yields a measure of direct discrimination. Our decomposition then yields a measure of systemic discrimination from the difference between total and direct discrimination. We find substantial total discrimination against Black applicants in managers' evaluations. These disparities are largely driven by systemic discrimination; while we find some evidence of direct discrimination, the majority of total discrimination is driven by race-based differences in prior work experience. Strikingly, the differences impact behavior despite evaluators being told that they were likely generated by direct discrimination elsewhere in the system. In light of prior work showing the effectiveness of information in reducing direct discrimination (Bohren et al. 2022), these findings highlight the difficulty of mitigating total discrimination when it is caused by systemic factors.

We organize the rest of this paper as follows. We next briefly review related literatures on systemic and direct discrimination. In Section 2 we present a simple motivating example that illustrates both forms. In Section 3 we develop our general formalization of direct and systemic discrimination, and in Section 4 we discuss mechanisms and present additional theoretical applications. Section 5 discusses identification, and develop our decomposition of total discrimination into direct and systemic components. Section 6 presents our lab experiment, and Section 7 presents our lab-in-the-field experiment. Section 8 concludes.

1.1 Related Literature

Our work builds on a large literature studying the role of systemic forces in driving group-based disparities (e.g., Pincus 1996; Feagin 2013; Allard and Small 2013; Pager and Shepherd 2008). While exact definitions vary, this systems-based approach distinguishes between direct discrimination, where individuals or firms treat people differently because of group identity itself, and indirect or systemic discrimination that considers the interlocking institutions or domains through which inequities propagate (Gynter 2003). In the systems-based approach, channels for observed disparities are taken as cumulative both within and across domains; discrimination is not just a product of a single individual or institution (Powell 2007). Systemic (or "structural") discrimination can be generated by the indirect relationships between outcomes and evaluations in roughly the same period, such as when discrimination in criminal justice drives unwarranted disparities in education and labor market outcomes.[6] It is also generated over time, such as when historic "redlining" practices in lending generates persistent disparities in credit access through its differential effects on generational wealth. The literature sometimes refers to the former as "side-effect" discrimination and the latter as "past-in-present" discrimination (Gynter 2003; Feagin and Feagin 1978; Feagin 2013).

Importantly, the systemic perspective shifts focus from the motives and biases of a given individual or institution to policies or institutional arrangements that contribute to *de facto* discrimination, perhaps without intent. Direct discrimination, either on the part of individuals or institutions, is inherently non-neutral: it arises from the explicit differential treatment of individuals on the basis of group identity. Systemic discrimination, in contrast, can exist in policies that are facially neutral by race, gender, or other protected characteristics (Hill 1988). For example, a lending algorithm which considers a person's zip code but does not use racial information when determining loan eligibility may be race neutral in design but discriminatory in practice. Black borrowers may be more likely to live in certain zip codes than equally creditworthy white borrowers, perhaps because of prior discriminatory policies in housing, employment, or financial markets (Aaronson, Hartley, and Mazumder 2021).[7]

The distinction between direct and indirect discrimination is echoed in legal theories of disparate treatment and disparate impact (e.g., Brekoulakis 2013; Gynter 2003; De Plevitz 2007; Rothstein 2017). Under the disparate impact doctrine, a policy or practice may be deemed discriminatory if it leads to disparities without substantial legitimate justification—as in *Griggs v. Duke Power Co. (1970)*.[8] A facially neutral practice may therefore be found to be discriminatory under this doctrine even in the absence of explicit categorization or animus. This notion of discrimination contrasts with the disparate treatment doctrine, which prohibits policies or practices motivated by a discriminatory purpose. Typically, proof of discriminatory intent is required for the finding of disparate treatment.[9]

A systemic perspective is also found in the recent literature on algorithmic unfairness

[6] Powell (2007) considers systemic discrimination as driving disparities within a domain, e.g., the hiring and promotion practices within a firm or industry, and structural discrimination as driving disparities through the interaction of different systems.

[7] Note that policies that are facially neutral on protected characteristics may not be neutral in intent. Mayhew (1968) argues that some organizations may have accepted Civil Rights legislation mandating "colorblind" treatment because they were aware systemic discrimination could preserve the status quo.

[8] See also *Dothard v. Rawlinson (1977)* and *Cocks v. Queensland (1994)*

[9] See, e.g., *Washington v. Davis (1976)* and *McClesky v. Kemp (1987)*.

(e.g., Angwin et al. 2016; Hardt, Price, and Srebro 2016; Zafar, Valera, Gomez Rodriguez, and Gummadi 2017; Berk, Heidari, Jabbari, Kearns, and Roth 2018; Kasy and Abebe 2021; Gebru 2020; Buolamwini 2022; Arnold, Dobbie, and Hull 2021). An algorithm which does not directly use protected characteristics may nevertheless return systematically disparate outcome predictions or treatment recommendations among equally qualified individuals. The literature studies how interlocking systems of data collection, model fitting, and human-algorithm decision-making may generate such disparities.

Finally, research in the field of stratification economics proposes a systemic perspective as necessary for understanding group-based disparities because advantaged groups have an incentive to maintain them (Darity 2005; Darity and Mason 1998; De Quidt, Haushofer, and Roth 2018). Without considering the systemic interactions generating a specific outcome, as well as the incentives involved in maintaining this system, a researcher or policy maker may miss important channels through which group-based disparities persist.

Our work also adds to the long literature on direct discrimination in economics, which is typically modeled as a causal effect of group membership on treatment.[10] Theoretical sources of direct discrimination include individual preferences or beliefs. In the canonical framework of taste-based discrimination, differential treatment emerges because individuals derive disutility from interacting with or providing services to members of a particular group (Becker 1957). In models of belief-based discrimination, differential treatment emerges because a decision-relevant statistic (such as labor market productivity) is unobserved, and there are group-based differences in beliefs about its distribution (Phelps 1972; Arrow 1973; Aigner and Cain 1977). While belief differences have traditionally been assumed to stem from true differences in the distributions, a recent literature has considered the role of inaccurate beliefs in driving direct discrimination (Bohren et al. 2022; Barron, Ditlmann, Gehrig, and Schweighofer-Kodritsch 2020; Hübert and Little 2020). These differences may stem from a lack of information or biased stereotypes (Bordalo, Coffman, Gennaioli, and Shleifer 2016; Coffman, Exley, and Niederle 2021; Bordalo et al. 2019; Fiske 1998), which again lead to causal effects of a protected characteristic on evaluations and decision-making.

A rich empirical literature in economics has largely followed this theoretical tradition. Research using both experimental and observational data has attempted to identify the causal effect of group identity on treatment, holding other observables constant (e.g., Bertrand and Mullainathan 2004; Fang and Moro 2011; Bertrand and Duflo 2016). In the widely-used correspondence study method, evaluators (e.g., hiring managers) are presented with information about individuals (e.g., applicants for a job), which consists of the individual's group

[10] Notable exceptions to the typical focus on direct discrimination in economics include Neal and Johnson (1996), Glover, Pallais, and Pariente (2017a), List (2004), Cook (2014), Hurst, Rubinstein, and Shimizu (2021), and Sarsons (2019). In Section 4.3 we discuss how the model of Coate and Loury (1993) captures a specific source of systemic discrimination in our framework.

identity and other signals of their qualifications (e.g., education level). Since everything but group identity—or a signal of this identity—is held constant in the experimental design, any differential treatment can be directly attributed to the causal effect of this variable. Recent advances in this methodology have been used to examine the dynamics of discrimination (Bohren et al. 2019) and the heterogeneity in discrimination across institutions (Kline, Rose, and Walters 2021).[11] A parallel empirical literature has developed tools to distinguish different economic theories of discrimination. Recent advances involve outcome-based tests of racial bias, in both observational (Knowles, Persico, and Todd 2001; Grau and Vergara 2021) and quasi-experimental data (Arnold, Dobbie, and Yang 2018; Hull 2021).

The systemic perspective suggests that standard economic tools for measuring direct discrimination misses an important component. Efforts to model and measure causation at any particular juncture and within a specific domain can substantially understate the cumulative impact of discrimination across domains or time. We contribute to the economics literature by expanding the tools for studying such forms of discrimination. Additionally, our framework offers new interpretations for previously documented group-based disparities. For example, evidence for a reversal of direct discrimination over time—such as the ones documented in Bohren et al. (2019) and Mengel, Sauermann, and Zölitz (2019)—may not imply that total discrimination has been mitigated or reversed. If, as argued, biased evaluators drive initial discrimination in the pipeline, the group that ends up being favored may still face substantial total discrimination when conditioning on underlying qualifications.[12]

2 A Motivating Example

We begin our analysis with a simple theoretical example that illustrates how systemic discrimination can emerge in a labor market. Suppose a platform matches workers and consumers to complete a task (e.g., an Uber ride, customer service query, or plumbing repair request). The worker completes the task for the consumer, after which the consumer assigns the worker a rating. After observing the rating, the platform selects a reward (e.g., a wage, continued employment, or increased future match propensity) for the worker.

Formally, the worker produces output of either low or high quality, $Q \in \{L, H\}$. The consumer observes this quality as well as the worker's race $G \in \{b, w\}$ before reporting a rating $R \in \{l, h\}$. Workers produce high quality with probability 1/2, independent of race.

[11] While Kline et al. (2021) refer to their study as estimating "systemic discrimination," this classification is not consistent with the large social science literature on systemic discrimination outlined above. Their correspondence study is designed to measure direct discrimination, formalized as the causal effects of protected characteristics in a hiring decision. We view this work as more accurately studying institutional direct discrimination, as formalized below.

[12] The systemic perspective also highlights the lasting impact of initial stereotypes (Bordalo et al. 2016, 2019). Even if signals become more precise and direct discrimination decreases, total discrimination can persist through systemic channels.

There are two types of consumers. With probability λ, a consumer is biased: he reports an accurate rating for white workers (i.e., h for quality H and l for quality L) but reports a rating of l for Black workers independent of their quality. With probability $1 - \lambda$, a consumer is unbiased and reports an accurate rating for workers from both groups. The biased consumers exhibit *direct* discrimination against Black workers, in that they rate Black workers who produce high quality output systematically lower than white workers producing the same output. This is the definition of discrimination most often used in economics: conditional on what a consumer observes (quality), she treats Black and white workers differently.

The platform does not have any inherent bias against Black workers: it seeks to offer a reward commensurate with output quality. If the platform observed quality itself, it would offer the worker a reward of \$1 for high quality and \$0 for low quality. However, since it instead relies on consumers' ratings, its prediction of quality (and thus reward) depends on both the realized ratings and its model of how consumers generate them. First suppose the platform fails to account for consumer bias: it takes a rating at face value and offers both Black and white workers a reward of \$1 for a high rating and \$0 for a low rating. This decision rule is prima facie race "neutral" in that it is the same for Black and white workers. The platform therefore does not exhibit direct discrimination because a white worker and a Black worker with the same rating are given the same reward.

However, conditional on actual output, a white worker who produces high quality receives an expected reward of \$1, while a Black worker who produces high quality receives an expected reward of $\$1 - \lambda$. This is because, conditional on high quality, the rating—and thus the reward—depends on the worker's race (while low quality results in a reward of \$0 independently of race). In turn, despite the platform treating all workers with the same rating equally, it generates race-based differences in outcomes for workers who producing output of identical quality.

This example motivates a broader notion of discrimination that captures disparities that are indirectly driven by group identity—namely, through the relationship between race and ratings. We refer to this indirect channel as *systemic discrimination.* Fixing the same reward rule (i.e., the mapping from ratings to rewards) for both groups, systemic discrimination corresponds to disparities arising from race-based differences in the *distribution* of rewards conditional on the same quality output. This is in contrast to direct discrimination, which stems from race-based differences in the reward rule itself, i.e., assigning a reward conditional on observing the same rating signal.[13] Here systemic discrimination by the platform is driven by direct discrimination by the customers, which results in the platform observing systematically higher ratings for white workers than Black workers with the same output.

[13] As we discuss further in Section 3.1, fixing the reward rule isolates systemic discrimination from any direct effects of worker group.

Our notion of *total discrimination* combines the direct and systemic channels. Formally, it corresponds to the difference between the reward distribution for white workers and Black workers who produce the same output. In other words, while systemic discrimination fixes the same reward rule for both groups, total discrimination does not—it measures disparities when each group is rewarded according to the reward rule for their group.

In the example thus far, total and systemic discrimination coincide because the platform uses the same reward rule for both groups; i.e., it does not exhibit direct discrimination. But this is not always the case. Suppose instead that the platform is aware of consumer bias and accounts for it when interpreting ratings: it offers a reward of $\$\lambda/(1+\lambda)$ to Black workers who receive a low rating, as this is the probability that they produced high output, and a reward of \$0 to white workers who receive a low rating, which perfectly signals low output. It continues to offer a reward of \$1 to all workers who receive a high rating as this perfectly conveys high output. In this case, the platform exhibits direct discrimination against white workers: conditional on a low rating, it offers a higher reward to a Black worker relative to a white worker. As before, consumers' direct discrimination still translates into systemic discrimination by the platform: fixing the reward rule for white workers (\$0 reward for a low rating and \$1 for a high rating), the average reward for white workers is 1/2 and the average reward for Black workers is $1/2(1-\lambda) < 1/2$).[14] However, the platform's direct discrimination in favor of Black workers exactly offsets the systemic discrimination against them, which eliminates total discrimination in the setting. That is, conditional on output, white and Black workers receive the same expected reward of 1/2.

From these two cases, we see that whether systemic discrimination translates into total discrimination depends crucially on whether the platform is aware of consumer bias: if it is unaware and takes the rating at face value, then it also exhibits total discrimination. In contrast, if it is aware, then it can engage in direct discrimination in the opposite direction to offset the systemic discrimination, eliminating total discrimination.

The simple model presented here highlights an important implication of broadening the definition of discrimination: when a signal is endogenously generated through the preferences and beliefs of an evaluator (e.g., the rating), then a subsequent evaluator can still exhibit systemic and total discrimination *even if* her own beliefs or preferences do not directly favor one group of workers. Given the rich literature in psychology and economics showing the inherent challenges of accurately predicting others' preferences and beliefs (Miller and McFarland 1987; Ross, Greene, and House 1977), or adjusting for biases in how a particular signal or outcome was generated even when the relevant information is provided (Andre 2022; Brownback and Kuhn 2019), it is plausible that initial biases or stereotypes will lead to

[14]Similarly fixing the reward rule for Black workers ($\$\frac{\lambda}{1+\lambda}$ reward for a low rating and \$1 for a high rating), the average reward for white workers ($\frac{1+2\lambda}{2(1+\lambda)}$), exceeds the average reward for Black workers, $\frac{1}{2}$).

persistent disparities despite prima facie race-neutral rules in subsequent evaluations. Measuring and accounting for systemic discrimination may be particularly important in settings where information is social—either because evaluators misperceive how other evaluators' make decisions, or because prior direct discrimination is baked into signals in a way that obscures its persistent impact.

The example also provides context for interpreting reversals of direct discrimination, as observed in recent work on dynamic discrimination (Bohren et al. 2019). Such reversals can belie persistent systemic and total discrimination against Black workers. For example, suppose the platform is only partially aware of the consumers' bias. Then consumers directly discriminate against Black workers while the platform reverses the direction of direct discrimination. However, Black workers will still face total discrimination in both stages.[15]

Importantly, bias in an initial evaluation is not necessary for social learning with "inflated" signals to lead to systemic discrimination. In Appendix B.1, we show how accurate statistical discrimination in an initial decision can also lead to persistent systemic discrimination. Social learning induces endogenous differences in the subsequent signaling technology, which drives this systemic discrimination; if the signaling technology were exogenous, such accurate statistical discrimination would not lead to systemic discrimination.

3 Formalizing Systemic Discrimination

We now develop a general theoretical framework to define systemic and total discrimination. This framework allows us to conceptually distinguish between direct discrimination, as typically considered in the economics literature, and the broader notions of discrimination considered in other fields. In the tradition of Becker (1957), Aigner and Cain (1977), and other classic analyses in economics, we develop this framework in the labor market context.

3.1 Setup

Consider a manager who evaluates a set of workers for a particular task. Each worker i has an observable group identity $G_i \in \{b, w\}$ and an *ex ante* unobservable productivity $Y_i^* \in \mathcal{Y}^* \subset \mathbb{R}$.[16] Before evaluating the worker, the manager observes a vector of k attributes $S_i \in \mathcal{S} \subset \mathbb{R}^k$ (e.g., educational background, prior evaluations, etc.). This vector provides a signal of productivity Y_i^*, potentially along with G_i. After observing G_i and S_i, the manager takes action $A_i \in \mathcal{A} \subset \mathbb{R}$. This action could be binary (e.g., whether or not worker i is hired for the task), continuous (e.g., the wage paid to worker i for completing the task), or something else (e.g., a multivalued rating). We abstract from complementarities across

[15] The example also highlights the sense in which "affirmative action"-type polices can mitigate systemic discrimination by inducing such reversals: loosening hiring thresholds for disadvantaged groups can serve the purpose of unwinding earlier discrimination without compromising expected productivity.

[16] We assume productivity and other relevant variables are real numbers to simplify notation; the analysis easily extends to more general sets.

workers and other realistic features of labor markets for simplicity; our analysis considers G_i, Y_i^*, S_i and A_i as *i.i.d.* across workers with some joint distribution.

Rather than explicitly modeling the manager's decision problem here, we take a reduced-form approach: the manager follows some systematic decision rule to determine her action choice from her information set. Formally we assume the existence of a function $A(g, s)$ that determines the manager's optimal action given a worker's group identity g and the signal s, such that $A_i = A(G_i, S_i)$. Absent restrictions on S_i, the existence of such rules is without conceptual loss. In Section 4 we provide a microfoundation for such rules as arising from a manager's preferences over (Y_i^*, G_i) and beliefs about the joint distribution of (Y_i^*, G_i, S_i).

To capture the idea that a worker's productivity in the task at hand can be affected by systemic forces (such as decisions made in other markets or time periods), we embed the hiring task in a larger economy. We assume worker i enters the economy with qualification $Y_i^0 \in \mathcal{Y}^0 \subset \mathbb{R}$. This captures a reference level of productivity; it could be the same as the payoff-relevant productivity in the hiring task, $Y_i^* = Y^0$, or Y_i^* could arise endogenously from Y_i^0 and the actions of other managers and firms. We consider Y_i^0 as an *i.i.d.* across workers and jointly distributed with (G_i, Y_i^*, S_i, A_i).

We do not explicitly model the relationship between Y_i^* and Y_i^0. Rather, we take Y_i^0 as a choice variable of the researcher. This choice allows us to formalize different notions of systemic discrimination within a unified framework, as we discuss below. We emphasize that Y_i^0 need not represent a fixed or "inherent" characteristic of the worker; it is a reference point for studying discrimination that emerges given initial conditions in a specific context. Note that setting Y_i^0 to a constant (i.e., $Y_i^0 = 0$) corresponds to the case where there are no initial qualification differences across protected groups.

Mapping this framework to our example from Section 2, quality Q corresponds to productivity, rating R corresponds to the signal, the reward corresponds to the action, and quality is taken to be the reference qualification. The following four non-employment contexts illustrate the generality of this setup.

Lending. A loan officer decides whether to lend to borrowers. Borrowers differ in their ability to pay back the loan (Y_i^*) if it is originated (A_i). Borrowers may differ in their initial lending qualifications (Y_i^0), which may interact with employment history and other factors to determine ability-to-repay. The loan officer observes borrowers' credit scores and income (S_i), which provide information about repayment ability.

Education. An admissions officer decides whether to admit students. Students differ in their academic performance (Y_i^*) if admitted (A_i). Students may differ in initial educational ability or motivation (Y_i^0), which may interact with prior educational opportunities and outside familial obligations to determine performance. The officer observes test scores and recommendation letters (S_i) which predict academic potential.

Healthcare. A doctor decides whether to test patients for a treatable disease. Patients differ in the disease outcome (Y_i^*) that is realized if they are not tested (A_i). The doctor observes blood pressure (S_i), which is informative about the disease state. Patients may differ in their underlying health (Y_i^0), which may interact with prior access to healthcare or time off from work to determine health outcomes.

Criminal Justice. A judge decides whether to release defendants before trial (A_i). Defendants differ in their potential for pretrial misconduct (Y_i^*) that is realized if they are released. Defendants may differ in their underlying propensity for criminal activity (Y_i^0), which interacts with access to basic necessities (e.g., transportation to return to court), employment opportunities, or other criminal justice conditions to determine the potential for pretrial misconduct. The judge observes defendants' prior criminal record (S_i), which provides information about pretrial misconduct potential.

In each context, one can imagine different ways in which qualification Y_i^0 interacts with decisions in other markets or domains to determine productivity Y_i^* by group G_i. Some of these differential interactions may arise from the kinds of direct discrimination typically considered in economics. The accumulation of such interactions across and within domains can lead to a broader notion of discrimination, as we next formalize.

3.2 Defining Direct, Systemic and Total Discrimination

Following Pincus (1996) and Gynter (2003), we delineate between two forms of discrimination in the manager's action with respect to worker group: *direct* and *systemic*. Direct discrimination arises causally from the worker's group identity itself, while systemic discrimination arises from group-based differences in the signal S_i which indirectly influences how actions depend on group identity. Such group-based differences in the signal may stem from direct discrimination in other periods or across different markets. Finally, *total discrimination* captures both direct and systemic forces. Direct, systemic, and total discrimination can occur at both the manager and firm level. We refer to discrimination by particular managers as *individual discrimination*, and, following Pincus (1996), refer to the aggregation of individual discrimination across managers as *institutional discrimination*. We present our definitions in the context of individual discrimination in this section, and show how they extend to the institutional level in Appendix C.

Formally, we define direct discrimination as group-based differences in manager actions, holding fixed the signal:

Definition 1 (Direct Discrimination)**.** *The manager exhibits direct discrimination at signal $s \in \mathcal{S}$ if $A(w, s) \neq A(b, s)$.*

Direct discrimination is a causal concept because it conditions on all relevant non-group characteristics S_i; it follows from the action rule's functional dependence on group membership. While Definition 1 considers direct discrimination at any signal realization s in the support of S_i, in practice researchers may focus on particular signal realizations or average over the signal distribution. Economic theory tends to focus on direct discrimination that stems from manager preferences or beliefs about productivity as a function of group identity.

Our definitions of total and systemic discrimination do not fix the signal, and hence incorporate non-causal elements of the relationship between manager actions and group membership. Total discrimination captures the overall relationship between the manager's action and the worker's group, holding fixed the workers' qualifications. Let $\mu^g(\cdot\,; y^0)$ denote the distribution of manager actions among workers of group g with qualification y^0 and let $\sigma^g(\cdot\,; y^0)$ denote the corresponding signal distribution conditional on y^0. The action distribution can be constructed from the signal distribution and action rule as $\mu^g(a; y^0) = \sigma^g(\{s : A(g, s) \in a\}; y^0)$ for any measurable set of actions $a \subset \mathcal{A}$. We say that total discrimination occurs if group w and b workers with the same qualification y^0 face different distributions over actions:

Definition 2 (Total Discrimination)**.** *The manager exhibits total discrimination at qualification y^0 if $\mu^b(a; y^0) \neq \mu^w(a; y^0)$ for some measurable set of actions $a \subset \mathcal{A}$, and the manager exhibits total discrimination if there exists a y^0 where this holds.*[17]

Fixing qualification, two components impact the action distribution for a worker: the action rule and the signal distribution. Differences in the action rules for workers from different groups will lead to different sets of signals mapping to a given action. For example, a Black worker may require more prior work experience than a white worker to receive a callback for the same job. Since the action likelihood depends on the set of signals that lead to this action, such differences in the action rules will translate to differences in the action distributions. This captures the direct discrimination component of total discrimination. The likelihood of an action also depends on the likelihood of the set of signals that lead to this action. Hence, group differences in the signal distributions for workers with a given qualification also lead to group differences in the action distributions. For example, white workers may be more likely to receive a callback than similarly qualified Black workers if white workers had greater access to internship opportunities and hence more work experience. This latter component is the channel for systemic discrimination, defined next. The first component can arise even when the signal distribution is the same for both groups; the second component can arise even when the action rule is the same for both groups.

To define systemic discrimination, we fix the group membership component of an action rule and consider how the action distribution for b versus w workers varies given their

[17] Equivalently, since $\mu^g(\cdot\,; y^0)$ gives the distribution of $A_i = A(G_i, S_i)$ conditional on $Y_i^0 = y^0$ and $G_i = g$, the manager exhibits total discrimination if $A(G_i, S_i)$ is not independent of G_i conditional on Y_i^0.

respective signal distributions. First fixing the action rule for group w, $A(w,s)$, we define a counterfactual action distribution for group b under this action rule, $\tilde{\mu}^b(a;y^0) \equiv \sigma^b(\{s : A(w,s) \in a\}; y^0)$ for any measurable set of actions $a \subset \mathcal{A}$. This captures the action distribution under the action rule for group w and the signal distribution for group b at qualification y^0. Comparing this counterfactual action distribution to the actual action distribution for group w determines whether group differences in the signal distribution translate to group differences in manager actions under the action rule for group w. Analogously, we can fix the action rule for group b and define the counterfactual action distribution for group w under this action rule, $\tilde{\mu}^w(a;y^0) \equiv \sigma^w(\{s : A(b,s) \in a\}; y^0)$ for any measurable set of actions $a \subset \mathcal{A}$. This yields the following definition:

Definition 3 (Systemic Discrimination)**.** *The manager exhibits systemic discrimination at qualification y^0 if $\mu^w(a;y^0) \neq \tilde{\mu}^b(a;y^0)$ or $\mu^b(a;y^0) \neq \tilde{\mu}^w(a;y^0)$ for some measurable set of actions $a \subset \mathcal{A}$, and the manager exhibits systemic discrimination if there exists a y^0 where this holds.*[18]

Since this definition fixes the action rule, systemic discrimination does not depend on the direct effect of group identity on manager actions, i.e., direct discrimination. Instead, it arises from the statistical relationship between the signal S_i and group identity. We condition this relationship on the qualification, such that systemic discrimination only arises among equally qualified workers with different signal distributions.

To illustrate these definitions, consider a simple example. Suppose all workers have the same qualification for a job and they receive a noisy performance evaluation (S_i) based on this qualification. Suppose managers promote male workers when they have an evaluation of 3 or higher on a 5 point scale, while female workers are promoted when they have an evaluation of 4 or higher. Moreover, bias in the evaluation process leads women to receive systematically lower evaluation scores than men. Our definition of direct discrimination captures the difference in evaluations required to achieve promotion: women have to generate a higher evaluation than men. Our definition of systemic discrimination instead compares the probability that male versus female workers achieve a particular evaluation, i.e., the difference in evaluation distributions conditional on qualification. Here, systemic discrimination captures disparities generated by women having lower evaluations than equally qualified men. Total discrimination aggregates both of these components: it compares the probability that a male worker generates an evaluation above 3 to the probability that a female worker generates an evaluation above 4. This captures the total difference in the probability of promotion for a male versus female worker.

[18]Equivalently, since $\tilde{\mu}^b(\cdot\,;y^0)$ gives the distribution of $A(w,S_i)$ conditional on $Y_i^0 = y^0$ and $G_i = b$ and $\mu^w(\cdot\,;y^0)$ gives the distribution of $A(w,S_i)$ conditional on $Y_i^0 = y^0$ and $G_i = w$, the manager exhibits systemic discrimination if $A(w,S_i)$ (or $A(b,S_i)$) is not independent of G_i conditional on Y_i^0.

Other examples that generate group-based differences in signals include word-of-mouth recruitment practices that prioritize workers with a social connection (S_i) to the firm. This can lead to systemic discrimination when men are more connected than equally qualified women (perhaps because of past direct discrimination in hiring). Similarly, the practice of "redlining" in mortgage markets can lead to borrowers from majority-white neighborhoods (as recorded in S_i) to be prioritized for a loan over borrowers from majority-Black neighborhoods. Such neighborhood-based prioritization can generate substantial race-based disparities despite the policy being prima facie race-neutral. If such treatment differences remain conditional on the relevant measure of qualification, then systemic discrimination will be present. In Section 4.4, we develop several additional examples to illustrate different ways systemic and total discrimination can arise.

Definition 3 aligns broadly with how systemic (or structural) discrimination is considered in the other literatures reviewed in Section 1.1: as a form of inequality operating indirectly through characteristics beyond group identity. The definition captures disparities that arise from the interaction between discriminatory decisions across time and domains. Systemic discrimination can emerge across time when past discriminatory decisions impact present decisions (so-called "past-in-present" discrimination, as illustrated in Section 2). Alternatively, current decisions can be impacted by anticipation of *future* discrimination (which we term "future-in-present" discrimination); for example, a Black defendant who anticipates a discriminatory jury may accept a less favorable plea agreement than a white defendant.[19] Systemic discrimination can emerge across domains when discriminatory practices in one market impact productivity or signaling in another (as illustrated in Section 4.4.2).[20] Finally, systemic discrimination can emerge when a system or institution is first "designed" by a group in power, which leads to the development of evaluation criteria that are optimized around the characteristics of this group.[21] In our framework, this corresponds to viewing the signal distribution as a choice variable for the dominant group. We emphasize systemic discrimination can generate total discrimination even when managers apply a group-neutral

[19]As a concrete example of ''future-in-present" systemic discrimination, Avery, Leibbrandt, and Vecci (2023) find that the use of artifical intelligence (AI) in recruitment increases the share of women applicants, as women anticipate less bias in recruitment when assessed by AI instead of humans.

[20]Powell (2007), for example, defines systemic discrimination as a "product of reciprocal and mutual interactions within and between institutions," both "within and across domains." He terms discrimination arising from the interactions of systems as "structural" and discrimination stemming from interactions in a system as "systemic." We do not formalize this distinction here, but it follows naturally from our framework.

[21]For example, De Plevitz (2007) discusses the impact of the "Eurocentric model of teaching" on schooling outcomes of Aboriginal children in Australia. She notes that by not accounting for the family structure and cultural obligations of the Aboriginal community, the educational system creates systemic barriers for the minority population. Similarly, the Australian Postal Commission required applicants to pass a medical examination including a height-to-weight threshold calibrated using Anglo-Saxon data, which led to the disproportionate rejection of South-East Asian applicants.

hiring rule because they fail to account for discrimination in other time periods or domains.[22]

3.3 The Choice of Y_i^0

Both systemic and total discrimination are defined with respect to the chosen measure of worker qualification Y_i^0, and are thus inherently tied to the researcher's choice of this reference point. At one extreme, when worker qualification is set equal to non-group characteristics observed by the manager ($Y_i^0 = S_i$), total discrimination is narrowly defined as any treatment disparities that remain when holding fixed the relevant observables. In this case, total and direct discrimination coincide and there is no role for systemic discrimination; this choice can thus be seen as implicit in most economic analyses of discrimination. At the other extreme, when worker qualification is set equal to a constant ($Y_i^0 = 0$), any unconditional group-based treatment disparity reflects (total) discrimination. This choice yields the broadest measure of systemic discrimination, which accounts for any indirect relationship between group identity and the payoffs or signals relevant to the present task.[23]

By selecting a Y_i^0 in between these two extremes, the researcher can bring focus to different systemic forces in the economy. When productivity in the hiring task depends on decisions in other markets or time periods, the researcher may wish to select an earlier measure of productivity as the reference qualification. For example, a worker's access to opportunity at university and subsequent employment history may impact her current labor market productivity Y_i^*. To consider the impact of employment history, the researcher can set Y_i^0 to be the worker's productivity when entering the labor market. In this case, total discrimination measures treatment differences in the present hiring task conditional on this initial labor market qualification. Alternatively, to account for both access to opportunity at university and employment history, a researcher could choose Y_i^0 to be a measure of human capital at matriculation to university. Both choices allow for the payoff-relevant outcome Y_i^* to depend on outside experiences (e.g., human capital accumulation). Systemic discrimination is especially important in this example, as by definition direct discrimination cannot capture endogenous disparities in the manager's payoff.

When non-group characteristics depend on decisions in other markets or time periods, the researcher may wish to fix the non-group characteristics observed in the outside decision as the reference qualification. For example, when a recruiter observes a worker's performance on a screening test and then makes a recommendation to a hiring manager, setting Y_i^0 to the

[22]For example, Pincus (1996) defines structural discrimination as referring to "the policies of dominant race/ethnic/gender institutions and the behavior of individuals who implement these policies and control these institutions, which are race/ethnic/gender neutral in intent but which have a differential and/or harmful effect on minority race/ethnic/gender groups." See also Hill (1988).

[23]See Rose (2022) for a related discussion in the case of direct discrimination. He argues that measuring discrimination—in his case, taste- or statistically-based—inherently requires taking a stance on what factors are decision-relevant for the evaluator, and what measures can be classified as discrimination.

screening test performance measures systemic discrimination in hiring manager actions that stems from direct discrimination by the recruiter (see Section 6.1 for an empirical example). Similarly, consider the case where racial, ethnic, or gender socialization affects the worker's decisions in a way that affects her work history or other manager signals (see Section 4.4.2 for a stylized example). To capture this channel as systemic discrimination, one can set Y_i^0 upstream of such socialization. Alternatively, one can allow for the possibility that workers of different groups have innately different preferences for certain job characteristics (e.g., schedule flexibility) by including measures of such preferences in Y_i^0.

Another focal case is setting Y_i^0 to the payoff-relevant outcome Y_i^*. In this case, total discrimination accounts for how workers from different groups with the same productivity for the task at hand are treated systematically differently. For example, suppose a training program or club membership serves solely as a signaling device and has no impact on the manager's payoff. A researcher may then wish to select a measure of discrimination that accounts for indirect discrimination stemming from differential access to the signaling opportunity.[24] Total discrimination with respect to qualification $Y_i^0 = Y_i^*$ encompasses this case, whereas direct discrimination does not.[25] This case aligns total discrimination with the legal notion of disparate impact, as it allows for disparities relevant to "business necessity."[26]

Thus, through the choice of Y_i^0, Definitions 1 to 3 provide a unified framework for studying different forms of direct, systemic, and total discrimination considered by various literatures. In any given setting, there may be one or several natural choices for Y_i^0 depending on which forms are of interest to the researcher. Our framework makes clear that the study of discrimination requires the researcher to take a normative stance on Y_i^0. While prior work often makes this choice implicitly, more explicit discussion may be critical for interpreting results and forming any appropriate policy response.

4 Sources of Discrimination

We now formalize potential sources of direct and systemic discrimination, as defined above. To do so, we first microfound the reduced-form action rule in terms of a manager's preferences and beliefs. We then delineate how the relationship between the signal, productivity,

[24] Note that this is the legal case sometimes made against group-based exclusivity in country clubs, which offer members a host of pecuniary and non-pecuniary benefits (Jolly-Ryan 1998).

[25] Alternatively, certain non-group characteristics may enter the manager's payoff in a way that is orthogonal to some objective measure of productivity, such as worker output. For example, a manager may have a preference for workers with shared alumni status or social connections even if these characteristics do not affect output. Setting Y_i^0 equal to the relevant measure of output allows the researcher to measure whether managers' preferences over non-group characteristics lead to systemic discrimination.

[26] Arnold et al. (2022), for example, consider a measure of disparate impact in the pretrial setting where $Y_i^0 = Y_i^*$ is a measure of pretrial misconduct potential. The $Y_i^0 = Y_i^*$ case also aligns total discrimination with some measures of algorithmic unfairness, in which A_i is a prediction of some latent state Y_i^* or an algorithmic recommendation based on such a prediction (Berk et al. 2018; Arnold et al. 2021).

and qualification can vary by group. The section ends with several additional theoretical applications to illustrate the different sources.

4.1 Setup

We first define the manager's decision problem. The manager's payoff depends on her action choice and the worker's productivity; it can also depend on the worker's group identity. Specifically, the manager receives payoff $u(a, y, g)$ from choosing action $a \in \mathcal{A}$ for a worker with productivity $y \in \mathcal{Y}^*$ and group $g \in \{b, w\}$. Since productivity is unobserved, the manager forms beliefs about its distribution from the signal and (potentially) the worker's group. We take a model misspecification approach and allow these beliefs to either be accurate or inaccurate (Bohren et al. 2022). Specifically, the manager believes productivity for group g has c.d.f. $\hat{F}^g(y)$ and the signal for a worker from group g with productivity y has distribution $\hat{\sigma}^g(\cdot\,; y)$. Given these distributions, the manager uses Bayes' rule to form a posterior c.d.f. $\hat{F}^g(y; s)$ of the worker's productivity after observing signal realization s. She chooses an action to maximize expected utility with respect to this posterior belief:

$$A(g, s) \equiv \arg\max_{a \in \mathcal{A}} \int_{\mathcal{Y}^*} u(a, y, g) d\hat{F}^g(y; s).$$

This provides a foundation for the decision rule defined in Section 3.1.

The *true* productivity distribution and signaling technology may also be relevant for the source of discrimination. Let $F^g(y; y^0)$ denote the productivity distribution for workers with qualification y^0 and group identity g. Let $\sigma^g(\cdot\,; y, y^0)$ denote the signaling distribution for workers with productivity y, qualification y^0, and group g. From these distributions, as well as the qualification c.d.f. $F_0^g(y^0)$, we can construct the true (unconditional) productivity distribution and signaling technology, denoted by $F^g(y)$ and $\sigma^g(\cdot\,; y)$, respectively. From Bayes' rule, we can analogously derive the posterior c.d.f. $F^g(y; s)$ of a worker's productivity conditional on observing signal realization s.

4.2 Sources of Direct Discrimination

Direct discrimination arises when the manager's action rule depends on group identity. This dependence stems from either the manager's preferences or beliefs. Only beliefs about the productivity distribution and signaling technology are relevant for the manager's decision rule; thus, these distributions are the relevant statistical sources of direct discrimination. In the case of classic (i.e., accurate) statistical discrimination, the manager has an accurate posterior belief about productivity that takes group membership into account, $\hat{F}^g(y; s) = F^g(y; s)$. There is direct discrimination when the posterior distribution depends on g, either because the productivity distribution $F^g(y)$ or the signaling technology $\sigma^g(s; y)$ differ by group (Phelps 1972; Arrow 1973; Aigner and Cain 1977). Direct discrimination can also arise from preferences or inaccurate beliefs, which is typically collectively referred to as "bias" in the economics literature. A canonical form of bias is taste-based discrimination, or animus, in which the manager's payoff $u(a, y, g)$ directly depends on group membership (Becker 1957). Another form of bias is inaccurate statistical discrimination (Bohren et al. 2022), in which the manager has an incorrect posterior belief about the worker's productivity, $\hat{F}^g(y; s) \neq F^g(y; s)$.[27]

4.3 Sources of Systemic Discrimination

Systemic discrimination arises from the interaction of two forces: how the manager's action rule depends on the signal, and how the signal depends on group identity and qualification. In contrast to direct discrimination, the *true* productivity distribution and signaling technology are relevant sources of systemic discrimination through their statistical relationship with qualification. Formally, systemic discrimination arises from the functional dependence of $A(g, s)$ on s—which is determined by the manager's preferences—and how the distribution $\sigma^g(s; y^0)$ depends on g. Since $\sigma^g(s; y^0)$ is constructed from the signaling technology and the productivity distribution conditional on qualification, i.e., $\sigma^g(s; y, y^0)$ and $F^g(y; y^0)$, there are two channels that can generate systemic discrimination: an informational channel given by group differences in $\sigma^g(s; y, y^0)$, and a technological channel given by group differences in $F^g(y; y^0)$. We discuss each channel in turn.

Informational Systemic Discrimination emerges from group-based differences in how signals are generated among workers who are equally productive at the task at hand and have the same qualification. Formally, it corresponds to the case where $\sigma^g(s; y, y^0)$ depends on g. Individuals may receive the same treatment conditional on the same signal realization, i.e., there is no direct discrimination, but conditional on Y_i^* and Y_i^0, the probability that worker i generates a given signal realization depends on her group g. For example, defendants with the same potential for pretrial misconduct (Y_i^*) and underlying propensity for criminal activity (Y_i^0) may have different likelihoods of a prior criminal offense (S_i) due to discrimination in policing. Or borrowers with the same ability to repay (Y_i^*) and initial lending qualification (Y_i^0) may have credit histories (S_i) that are differentially informative due to discrimination in past borrowing opportunities (Bartik and Nelson 2016).

One focal form of informational systemic discrimination is *signal inflation*, in which a component of S_i is systematically higher for one group than the other. When higher signal realizations lead to more favorable actions, this generates systemic discrimination against

[27] Such inaccurate beliefs can arise from biased stereotypes (Bordalo et al. 2016), self-image concerns (Bohren and Hauser 2022; Barron et al. 2020), or limited attention (Bartoš, Bauer, Chytilová, and Matějka 2016). Bias can also stem from the manager accurately predicting and acting on a non-productive outcome $\tilde{Y}_i$ e.g., the manager's payoff depends on $\tilde{Y}_i \neq Y_i^*$. The computer science literature sometimes refers to this channel as "omitted payoff bias" (Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018); see also Canay, Mogstad, and Mountjoy (2020) and Grau and Vergara (2021) for discussions of this issue.

the group with the lower average signal. In the previous criminal justice example, suppose white defendents are more likely to have no prior criminal records than Black defendents, and that a clean record increases the probability of being released on bail (Pager, Bonikowski, and Western 2009; Agan and Starr 2017b). Such signal inflation might arise because, for example, S_i is affected by direct discrimination in an earlier period or a separate domain, e.g., Black individuals may be more likely to be stopped by police (Pierson, Simoiu, Overgoor, Corbett-Davies, Jenson, Shoemaker, Ramachandran, Barghouty, Phillips, Shroff et al. 2020).

Social information—that is, signals that correspond to other managers' actions—and inaccurate beliefs about how this information is generated are key for signal inflation in perpetuating disparities. In the criminal justice example, suppose the bail judge believes that there is no direct discrimination in policing, and therefore, a prior criminal offense reflects the same underlying propensity for criminal activity for both racial groups. In reality, there may be no underlying race-based difference in criminal activity: the differential likelihood of having a clean criminal record stems from direct discrimination in prior policing decisions. Inaccurate beliefs about policing will then lead to systemic and total discrimination through signal inflation. This channel is illustrated in the motivating example (Section 2) and empirically documented below.

Another focal form of informational systemic discrimination is *screening discrimination*, where the manager has a more precise (i.e., lower variance) signal for one group than the other. Observing the signal then leads to a larger reduction in uncertainty over productivity for the group with the more informative signal. Unlike signal inflation, the direction of systemic discrimination from screening has a heterogeneous impact depending on the worker's qualification. Consider, for example, a binary hiring decision in which the signal is normalized to be expected productivity and the worker's qualification corresponds to realized productivity. Then higher signal variance benefits low productivity workers, as it leads to more workers realizing signals above the hiring threshold. In contrast, it is detrimental to high productivity workers, as it leads to more workers realizing signals below the hiring threshold. Differences in precision may arise, for example, when the signal is a test specifically trained to screen group-w workers and generates less reliable information about the productivity of group-b workers.[28] Group-b may also have less informative signals because of fewer previous opportunities to establish a record, as in the credit example discussed above. We illustrate this channel in Section 4.4.1 and also document it empirically below.

Technological Systemic Discrimination emerges from group-based differences in productivity Y_i^* for workers with the same initial qualification Y_i^0. Formally, it is generated

[28] This case was documented in recent work showing that subjective tests designed to screen men led to disparate outcomes for women; amending or replacing the tests with more objective evaluations mitigated disparities (Mocanu 2022). De Plevitz (2007) similarly documents systemic discrimination due to the use of height-to-weight ratios calibrated with Anglo-Celtic data in job screening.

when $F^g(y; y^0)$ depends on g. This channel is clearly only possible when the chosen qualification measure differs from productivity in the current task, $Y_i^0 \neq Y_i^*$. Here, there can be systemic discrimination even when the signaling technology conditional on productivity is identical across groups. Similar to informational systemic discrimination, this technological channel can take the form of inflated productivity, in which Y_i^* is systematically higher for group-w workers relative to group-b workers with the same initial qualification. For example, this could arise if group-w workers have more access to training and skill development than group-b workers.[29] Technological systemic discrimination can also arise from other properties of the productivity distribution, fixing initial qualification. For example, differential selection into and exit from prior tasks may impact the productivity distribution of the workers who remain in the market for the current task.[30]

The model in Coate and Loury (1993) generates one form of technological systemic discrimination, due to anticipation of future direct discrimination. In this model, workers start with no group-based differences in initial qualification (Y_i^0) and make a costly decision to invest in human capital that increases their productivity for a particular task (Y_i^*). A manager observes the workers' group membership and a signal of productivity, but not the investment decision, before deciding whether to hire the worker for a better paying job that requires higher productivity to succeed. The optimal hiring decision depends on the employer's beliefs about whether w-workers are more likely to invest in human capital than b-workers, or vice versa. Beliefs are self-fulfilling in the sense that if b-workers believe that employers are less likely to hire them for the higher paying job than w-workers, they are less likely to invest in human capital; employers' beliefs about group-based differences in productivity are then correct in equilibrium. This generates group-based differences in Y_i^* conditional on initial qualification Y_i^0. Accurate statistical direct discrimination arises in hiring because workers from different groups receive differential treatment due to these actual group-based differences in the productivity distribution. However, since higher productivity leads to higher signals, the endogenous difference in the productivity distributions also translates into different signal distributions for group-b and group-w workers relative to the same pre-investment qualification, leading to systemic discrimination against group-b workers.

[29] Gallen and Wasserman (2021) highlight this channel when documenting gender differences in career advice. There, women seeking information about professional opportunities are more likely to receive advice about work/life balance than similar requests by men. The authors argue that this can deter investment in human capital and the pursuit of careers in competitive fields.

[30] Analogous to how direct discrimination can arise from omitted payoff bias (see Footnote 27), technological systemic discrimination can arise when the manager's payoff depends on non-group characteristics that do not directly impact the firm-relevant measure of productivity. Such a characteristic may be observable, and hence a component of S_i, or unobservable and predicted by S_i. For example, a manager may have a preference for workers with shared alumni status or social connections, even if these characteristics do not affect output. Indeed, Glover, Pallais, and Pariente (2017b) find evidence for such manager-worker homophily as a source of discrimination.

Technological systemic discrimination also includes the type of "task-based" discrimination studied in Hurst et al. (2021). In their model, workers start out with no group-based differences in initial qualification (Y_i^0), and then proceed to specialize in tasks that vary along a number of attributes, including the extent to which they require interaction with others. Firms pay different wages depending on a workers specialization on a given task (Y_i^*). Importantly, although there are no group-based differences in initial qualifications, racial barriers to specialization can generate group-based differences in Y_i^*.

Importantly, in our framework, group differences in the distribution of worker qualification cannot lead to systemic discrimination with respect to that qualification because the definition of systemic discrimination conditions on qualification. This observation highlights how the chosen qualification measure is an important reference point: only disparities that emerge subsequent to it contribute to systemic discrimination with respect to it. At the one extreme, when Y_i^0 is set to a constant, all differences in the unconditional signaling technology $\sigma^g(s;y)$ and the unconditional productivity distribution $F^g(y)$ contribute to systemic discrimination. At the other extreme, when Y_i^0 is set to Y_i^*, only differences in the unconditional signaling technology $\sigma^g(s;y)$ contribute to systemic discrimination: differences in $F^g(y)$ play no role. In between these extremes, differences in the conditional signaling technology $\sigma^g(s;y,y^0)$ and the conditional productivity distribution $F^g(y;y^0)$ can both contribute to systemic discrimination. We also note there is no scope for "inaccurate" systemic discrimination: only true distributions contribute to systemic discrimination.[31]

Given that accurate statistical (direct) discrimination also stems from group-based differences in the signal and productivity distributions, it is instructive to highlight how it conceptually differs from the sources of systemic discrimination. Accurate statistical discrimination arises from the impact of these distributions on the action rule; in contrast, systemic discrimination arises from the impact of these distributions on the action *distribution*. When $Y_i^0 = Y_i^*$, differences in the signaling distribution can lead to both informational systemic discrimination and accurate statistical direct discrimination, while differences in the productivity distribution can only lead to the latter. Focusing only on direct discrimination would miss a key aspect of how group differences in the signal distribution contribute to action disparities. When $Y_i^0 \neq Y_i^*$, differences in the productivity distribution conditional on qualification can also lead to technological systemic discrimination; again, focusing only on direct discrimination would miss a key driver of disparities.

[31]Inaccurate beliefs about $\sigma^g(s;y,y^0)$ could, however, lead to inaccurate perceptions about the extent to which different signaling technologies lead to systemic discrimination, and therefore the choice of which signaling technology to use if one seeks to avoid systemic discrimination. For example, a mortgage assessor may perceive the signaling technology for a particular credit score to be identical across groups, and therefore, choose to continue using it despite discriminatory signal inflation.

4.4 Additional Examples

We saw in Section 2 how bias in an initial evaluation can lead to systemic discrimination in subsequent evaluations through signal inflation. We now present two additional examples to illustrate other sources of systemic discrimination: screening and direct discrimination in a concurrent decision in another domain.

4.4.1 Systemic Discrimination in Worker Screening

Overview. This example shows how group-based differences in the precision of productivity signals can lead to both direct and systemic discrimination in a screening action. The former channel is through accurate statistical discrimination: the groups face different effective thresholds for the same signal realizations because of the difference in signal precision. The latter systemic channel comes from the difference in the signal distribution, accounting for the difference in thresholds. For example, if an aptitude test is designed by a dominant group it may provide more accurate information about members of that group than for a minority group; alternatively, a medical diagnostic test may only be trialed on the majority group and is thus more predictive for this group. Such disparities in screening accuracy corresponds to a type of systemic discrimination: even if individuals from different groups receive the same treatment conditional on the same test result, if the system neglects developing accurate methods to screen minority groups these groups will face systemic discrimination.

Our example shows how canonical statistical discrimination models may not capture the full extent of (total) discrimination stemming from differences in the signaling technology. It also shows how discrimination due to differences in the signaling technology manifests in fundamentally different ways than discrimination due to differences in the prior distribution of productivity (i.e., the other source of classic statistical discrimination). When the qualification is set to current productivity, $Y_i^0 = Y_i^*$, the former can lead to both direct and systemic forms of discrimination in the current decision, while the latter only leads to direct discrimination (as illustrated in Appendix B.1). Finally, this example shows how systemic discrimination from disparities in the informativeness of signals is likely to be heterogeneous across worker productivity levels: more productive workers tend to face more systemic discrimination than less productive workers.

Application. Suppose worker productivity is distributed identically within groups, $Y_i^* \mid G_i \sim N(0,1)$, but the manager's signal $S_i = Y_i^* + \varepsilon_i$ has a group-specific precision: $\varepsilon_i \mid G_i \sim N(0, 1/\eta_{G_i})$, with more precise signals for group-w, $\eta_w > \eta_b > 0$. The distribution of S_i for a group-g worker with productivity y is $N(y, 1/\eta_g)$ and the posterior expected productivity for a worker from group g who generates signal realization s is $s\eta_g/(1+\eta_g)$.

Suppose the manager hires all workers whose posterior expected productivity is at or above some threshold $t \in \mathbb{R}$: $A(g,s) = \mathbb{1}\{s\eta_g/(1+\eta_g) \geq t\}$. The manager thus hires group-g workers with signal realizations $S_i \geq t(1+\eta_g)/\eta_g$. Group-$b$ workers face a higher signal threshold, since $(1+\eta_b)/\eta_b > (1+\eta_w)/\eta_w$. Therefore, there is direct discrimination against group b stemming from the higher cutoff arising from their less precise productivity signal. Specifically, group-w workers with $S_i \in (t\frac{1+\eta_w}{\eta_w}, t\frac{1+\eta_b}{\eta_b}]$ are hired but group-b workers with signals in this range are not (hiring of workers with other signals does not depend on group).

Even without the direct discrimination in signal thresholds, however, the difference in signal precision causes equally-productive workers to be hired at different rates depending on their group. For a given $y \in \mathcal{Y}$ and $g \in \{b, w\}$, systemic discrimination is captured by

$$\begin{aligned}
&E[A(g,S_i)|Y_i^* = y, G_i = w] - E[A(g,S_i)|Y_i^* = y, G_i = b] \\
&= Pr\left(S_i \geq t(1+\eta_g)/\eta_g | Y_i^* = y, G_i = w\right) - Pr\left(S_i \geq t(1+\eta_g)/\eta_g | Y_i^* = y, G_i = b\right) \\
&= \Phi\left(\eta_b\left(t(1+\eta_g)/\eta_g - y\right)\right) - \Phi\left(\eta_w\left(t(1+\eta_g)/\eta_g - y\right)\right),
\end{aligned}$$

where $\Phi(\cdot)$ gives the standard normal distribution.[32] Since $\eta_b \neq \eta_w$, this expression is non-zero unless $y = t\frac{1+\eta_g}{\eta_g}$. Therefore, there is systemic discrimination almost everywhere in the productivity distribution, stemming from the differential probabilities of the signal being above a given cutoff for equally productive group-w versus group-b workers.

Systemic discrimination in this screening action is heterogeneous across worker productivity levels. With $\eta_w > \eta_b > 0$, the systemic discrimination hurts group-b workers at high levels of productivity (where $y > t\frac{1+\eta_g}{\eta_g}$) and *favors* group-$b$ workers at low levels of productivity (where $y < t\frac{1+\eta_g}{\eta_g}$) since $\Phi(\cdot)$ is strictly increasing. Intuitively, having a higher signal variance makes low-productivity group-b workers more likely to have a signal above the effective threshold by chance, while high-productivity group-b workers are more likely to generate a signal below the threshold by chance.

The average level of systemic discrimination across workers depends on which of these two productivity groups is larger. In a "cherry-picking" market with $t > 0$, such that a minority of workers are hired in each group (i.e., $Pr\left(S_i \geq t\frac{1+\eta_g}{\eta_g} \mid G_i = g\right) < 0.5$), the systemic discrimination favors group-b overall. Here, there are fewer high-productivity group-b workers hurt by the higher signal variance than low-productivity group-b workers helped by it. Conversely, in a "lemon-dropping" market with a majority of workers hired ($t < 0$) the systemic discrimination hurts group-b workers overall.

This application highlights the issue of examining screening discrimination using only direct measures, as this will miss an important component of how differential signal precision impacts total discrimination in the setting.

[32]For the second equality, we use the fact that $\eta_g(S_i - y) \mid \{Y_i^* = y, G_i = g\} \sim N(0,1)$ so $Pr(S_i \geq t\frac{1+\eta_g}{\eta_g}|Y_i^* = y, G_i = g') = Pr(\eta_{g'}(S_i - y) \geq \eta_{g'}(t\frac{1+\eta_g}{\eta_g} - y)|Y_i^* = y, G_i = g') = 1 - \Phi(\eta_{g'}(t\frac{1+\eta_g}{\eta_g} - y))$.

4.4.2 Signaling Across Markets

Overview. This example shows how direct discrimination in one market can lead to systemic discrimination in another market through endogenous worker investments in the signaling technology. It highlights that systemic discrimination need not be dynamic: it can emerge through the contemporaneous interactions in treatment between markets or domains—what Feagin and Feagin (1978) call "side-effect" discrimination. We base this example on the field experiment of Bursztyn, Fujiwara, and Pallais (2017), where single women were found to report lower desired salaries and less preference for workplace flexibility when they expected peers to see their reports of these traits. This example also speaks to socialization as a potential mechanism for informational systemic discrimination, where seemingly inherent traits (such as "competitiveness" or "assertiveness") are expressed differentially among equally qualified individuals as a function of group identity in order to influence other objectives.

Application. Suppose a worker's choice of a trait S_i is observed and used to assess the payoff-relevant outcome in two markets: the job market and the marriage market. Each worker i from group $G_i \in \{m, f\}$ has an initial level $Y_i^* \in \mathbb{R}$ of the trait, which can be viewed as her "natural" or "endowed" level before any action can be taken to alter it. For a private cost, the worker can then take actions that either raise or lower the observable level of her trait. In other words, the worker strategically chooses $S_i \in \mathbb{R}$ given Y_i^*. Suppose the cost to alter S_i away from Y_i^* is quadratic in the distance between the chosen and endowed trait: to set $S_i = s$ when $Y_i^* = y$ the worker bears a cost of $C(s, y) = (s - y)^2$.

Evaluators differentially value the outcome that the trait signals across the two markets. Suppose evaluators are unaware of the workers' ability to distort their signal, and believe $E[Y_i^* \mid S_i = s] = s$ as in the setup of Section 2. In the job market, recruiters prefer higher levels of Y_i^* for both groups and have a common action rule of $A_1(g, s) = s$. In the marriage market, prospective partners prefer higher levels of the trait among workers of group-m and lower levels of the trait among workers of group-f. Partner actions in this market are given by $A_2(m, s) = s$ and $A_2(f, s) = -s$. There is thus no direct discrimination in the job market, but there is preference-based (direct) discrimination in the marriage market.

Workers value the chosen action in each market, with weight $\gamma \in [0, 1]$ on the job market action and $1 - \gamma$ on the marriage market action. A worker from group g with an endowed trait level of y therefore chooses $S_i = S(G_i, Y_i^*)$, where

$$S(g, y) \equiv \arg\max_{s \in \mathbb{R}} \gamma A_1(g, s) + (1 - \gamma) A_2(g, s) - (s - y)^2.$$

For group-m, this leads to an endogenously inflated signal: $S(m, y) = y + \frac{1}{2} > y$. Whether or not group-$f$ workers inflate their signal depends on whether they put more weight on the job or marriage market: $S(f, y) = y + \gamma - \frac{1}{2}$. Intuitively, when $\gamma > \frac{1}{2}$ the labor market benefit of a small increase in S_i from the endowed Y_i^* is larger than the marginal cost of such inflation on the marriage market: $S(f, y) > y$. But when $\gamma < \frac{1}{2}$ the marriage market penalty induces the worker to shade down her endowed trait, with $S(f, y) < y$. Note that in the extreme case of $\gamma = 1$ the two groups have identical choices of $S(g, y) = y + \frac{1}{2}$, as the marriage market discrimination has no effect on group-f's choices in this case.

When $\gamma \neq 1$, such that the marriage market affects the signal choice of group-f workers, there is systemic discrimination in the job market. Setting $Y_i^0 = Y_i^*$, we have $E[A_1(g, S_i)|Y_i^0 = y, G_i = m] - E[A_1(g, S_i)|Y_i^0 = y, G_i = f] = 1 - \gamma > 0$.[33] Intuitively, the direct discrimination group-f workers face on the marriage market causes them to invest differently in the signaling technology than equally productive group-m workers. Since there is no direct discrimination, $A_1(m, s) = A_1(f, s)$, total discrimination is entirely driven by this channel. A conventional analysis that conditions on or randomizes over the endogenous signals to measure direct discrimination would thus fail to detect discrimination in this setting.

5 Measuring Systemic Discrimination

We now develop measures of systemic discrimination which leverage novel decompositions of total discrimination into direct and systemic components. We first present these decompositions before discussing the identification of each component.

For notational simplicity, we assume throughout this section that actions are real-valued (i.e., $\mathcal{A} \subset \mathbb{R}$) and focus on measures of discrimination that correspond to mean differences by group.[34] Total discrimination at qualification level $y^0 \in \mathcal{Y}^0$ is given by:

$$\Delta(y^0) \equiv E[A(G_i, S_i) \mid G_i = w, Y_i^0 = y^0] - E[A(G_i, S_i) \mid G_i = b, Y_i^0 = y^0]. \quad (1)$$

A finding of $\Delta(y^0) > 0$, for example, would mean that group-w workers with qualification y^0 are hired at a higher rate than equally-qualified group-b workers. Correspondingly, direct discrimination at signal realization $s \in \mathcal{S}$ is given by:

$$\tau(s) \equiv A(w, s) - A(b, s), \quad (2)$$

A finding of $\tau(s) > 0$, for example, would mean, that belonging to group w vs. b causes workers with signal s to be hired more often. Finally, systemic discrimination at qualification level $y^0 \in \mathcal{Y}^0$ is given by:

$$\delta(g, y^0) \equiv E[A(g, S_i) \mid G_i = w, Y_i^0 = y^0] - E[A(g, S_i) \mid G_i = b, Y_i^0 = y^0], \quad (3)$$

[33] There is also systemic discrimination in the marriage market: $E[A_2(g, S_i)|Y_i^0 = y, G_i = m] - E[A_2(g, S_i)|Y_i^0 = y, G_i = f]$ equals $1 - \gamma$ for $g = m$ and $\gamma - 1$ for $g = f$.

[34] This analysis of means easily generalizes to other distributional features of A_{ij}, such as variances or higher-order moments. For a complete distributional analysis one could consider mean disparities in the indicators $\mathbb{1}[A_i \leq a]$ for $a \in \mathcal{A}$.

for $g \in \{w, b\}$. A finding of $\delta(g, y) > 0$, for example, would capture higher hiring rates among equally-productive group-w vs. group-b workers that arises indirectly from the signal.

5.1 Decomposing Total Discrimination

Our decomposition of total discrimination into direct and systemic components follows directly from Equations (1) to (3):

$$\overbrace{\Delta(y^0)}^{\text{Total discrimination}} = \underbrace{E[\tau(S_i) \mid G_i = w, Y_i^0 = y^0]}_{\text{Average direct discrimination}} + \underbrace{\delta(b, y^0)}_{\text{Systemic discrimination}}, \quad (4)$$

where we add and subtract $E[A(b, S_i) \mid G_i = w, Y_i^0 = y^0]$ to and from the definition of $\Delta(y^0)$ and rearrange terms.[35] Equation (4) shows that total discrimination at qualification level y^0 can be written as the sum of two terms: average direct discrimination across the signal space, where the average is taken with respect to the signal distribution for workers from group w with qualification level y^0, and systemic discrimination at qualification level y^0 when the manager uses the action rule for group b.

Equation (4) is in the spirit of Kitagawa (1955), Oaxaca (1973), and Blinder (1973), who relate unconditional disparities to a component explained by observable worker characteristics (e.g., education or labor market experience) and a residual "unexplained" disparity. These classic decompositions can be viewed as a strategy for measuring direct discrimination, which attempts to hold fixed all relevant non-group characteristics. Equation (4), in contrast, leads to strategies for measuring systemic discrimination by the residual of total description after accounting for direct discrimination—which we develop further below.

As with the classic Kitagawa-Oaxaca-Blinder approach, there are multiple equivalent ways to decompose total discrimination into direct and systemic components and the "order" of the decomposition may matter empirically. In particular, we also have

$$\Delta(y^0) = E[\tau(S_i) \mid G_i = b, Y_i^0 = y^0] + \delta(w, y^0) \quad (5)$$

by adding and subtracting $E[A(w, S_i) \mid G_i = b, Y_i^0 = y^0]$ to and from the definition of $\Delta(y^0)$ and rearranging terms. Equation (5) decomposes total discrimination into average direct discrimination with respect to the signal distribution for workers from group b and systemic discrimination when the firm uses the action rule for group w, all at qualification level y^0.

[35] Specifically, $\Delta(y^0) = E[A(w, S_i) \mid G_i = w, Y_i^0 = y^0] - E[A(b, S_i) \mid G_i = b, Y_i^0 = y^0] - E[A(b, S_i) \mid G_i = w, Y_i^0 = y^0] + E[A(b, S_i) \mid G_i = w, Y_i^0 = y^0] = E[A(w, S_i) - A(b, S_i) \mid G_i = w, Y_i^0 = y^0] + (E[A(b, S_i) \mid G_i = w, Y_i^0 = y^0] - E[A(b, S_i) \mid G_i = b, Y_i^0 = y^0])$. The first expectation equals $E[\tau(S_i) \mid G_i = w, Y_i^0 = y^0]$ while the second term in parentheses equals $\delta(b, y^0)$.

Averaging Equations (4) and (5) yields a third decomposition:

$$\Delta(y^0) = \bar{\tau}(y^0) + \bar{\delta}(y^0), \tag{6}$$

where $\bar{\tau}(y^0) \equiv \frac{1}{2}(E[\tau(S_i) \mid G_i = w, Y_i^0 = y^0] + E[\tau(S_i) \mid G_i = b, Y_i^0 = y^0])$ is an unweighted average of the direct discrimination terms in equations Equations (4) and (5), while $\bar{\delta}(y^0) \equiv \frac{1}{2}(\delta(w, y^0) + \delta(b, y^0))$ is an unweighted average of the systemic discrimination terms.

Each of the three decompositions (4)-(6) yield a measure of systemic discrimination, given by the difference between total discrimination and the direct discrimination component. The challenge of identifying systemic discrimination thus reduces to the challenge of measuring direct and total discrimination. We next discuss how identification can be achieved, starting with the case where the researcher-chosen qualification metric Y_i^0 is observed.

5.2 Observable Y_i^0: The Iterated Audit Design

When worker qualification is directly observed, it can be conditioned on to identify total discrimination: $\Delta(y^0) = E[A_i \mid G_i = w, Y_i^0 = y^0] - E[A_i \mid G_i = b, Y_i^0 = y^0]$ for each $y^0 \in \mathcal{Y}^0$. Qualification may be observed when it is chosen to be a simple predetermined characteristic, such as a worker's educational attainment prior to joining the labor market. In the case of $Y_i^0 = 0$, i.e., when the researcher sets qualification as constant across workers, total discrimination is identified by the unconditional disparity $E[A_i \mid G_i = w] - E[A_i \mid G_i = b]$.

When Y_i^0 is observed, a simple experimental approach can identify the direct and total discrimination components in equations (4)-(6). We term this approach an *iterated audit* (IA), as it applies tools from conventional audit or correspondence studies in multiple stages to empirically separate direct and systemic discrimination.

The first IA step randomizes manager perceptions of group membership, as in a conventional audit or correspondence study, among a real set of workers with a given qualification level. Formally, in a population of workers with a given distribution of (G_i, S_i, Y_i^*, Y_i^0), the researcher generates a $\tilde{G}_i$ such that manager actions are given by $A_i = A(\tilde{G}_i, S_i)$ and where $\tilde{G}_i \perp\!\!\!\perp (G_i, S_i, Y_i^*) \mid Y_i^0$ by virtue of the randomization. This $\tilde{G}_i$ "treatment" can be used to measure the causal effect of group identity. For example, a researcher may take a set of real white and Black resumes and randomize distinctively white and Black names among equally-qualified workers, holding fixed all other information on the resume.[36] The researcher then elicits manager actions A_i in the experimental sample. Comparing the response to group-w

[36]We abstract away from several conceptual issues with measuring direct discrimination by manipulating signals of group membership, such as worker names, instead of the perceived characteristic directly. Such issues can be especially important when G_i is meant to capture race. See, e.g., Fryer and Levitt (2004); Sen and Wasow (2016); Gaddis (2017); Kohler-Hausmann (2019) for discussions of these issues. Notably, Rose (2022) develops a theoretical framework demonstrating the issues present with inferring perceived social identity from race as coded in the specific datasets. This coding can create issues with measurement error and interpretation of disparities as direct discrimination by animus versus statistical discrimination.

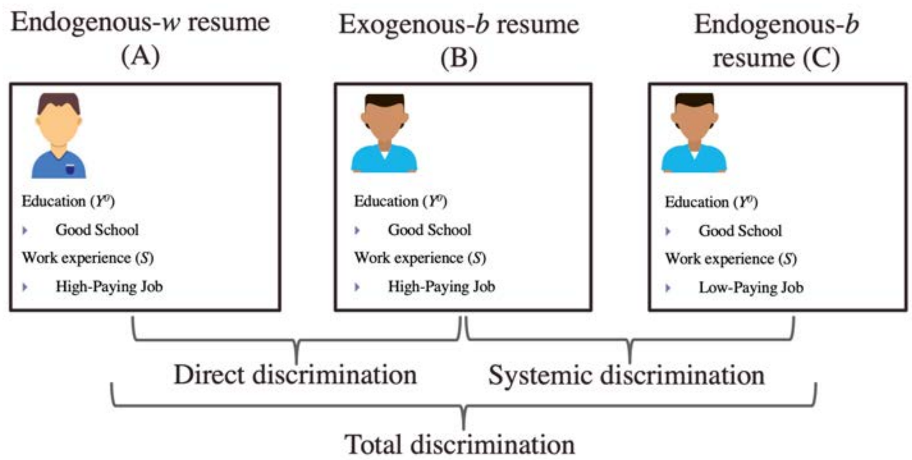


Figure 1. Iterated Audit Example

workers randomized to $\tilde{G}_i = w$ with the response to group-w workers randomized to $\tilde{G}_i = b$, at qualification level y^0, identifies the direct discrimination component of (4):

$$E[A_i \mid \tilde{G}_i = w, G_i = w, Y_i^0 = y^0] - E[A_i \mid \tilde{G}_i = b, G_i = w, Y_i^0 = y^0] = E[\tau(S_i) \mid G_i = w, Y_i^0 = y^0],$$

Similarly, comparing the response to group-b workers randomized to $\tilde{G}_i = w$ with the response to group-b workers randomized to $\tilde{G}_i = b$, at qualification level y^0, identifies the direct discrimination component of (5). Averaging these comparisons identifies the direct discrimination component of (6).

The second IA step measures *total discrimination* by eliciting manager actions among workers whose perceived group membership was not manipulated by the experiment. This could be in a separate non-experimental sample, with the same distribution of (G_i, S_i, Y_i^*, Y_i^0), or among workers with $G_i = \tilde{G}_i$ in the experimental sample. Subtracting one of the three direct discrimination components estimated in the first step from the total discrimination measure $\Delta(y^0) = E[A_i \mid G_i = w, Y_i^0 = y^0] - E[A_i \mid G_i = b, Y_i^0 = y^0]$ identifies one of the three systemic discrimination components in equations (4)-(6).

Figure 1 illustrates an example iterated audit conducted with white (group w) and Black (group b) resumes, where the researcher is interested in studying discrimination conditional on worker education Y_i^0. Resumes A and C represent "endogenous" profiles of white and Black applicants with the same level of education. Disparities in hiring decisions (such as callback rates) between such resumes capture total discrimination given this choice of Y_i^0. Resume b represents an "exogenous" profile of a Black candidate with information matching the white candidate's, generated by randomizing a distinctively Black photo to the real white resume A (implicitly holding fixed all other elements, such as education and work experience). Hiring disparities arising from this randomized "treatment" capture direct discrimination, as in a classic correspondence study. The residual contrast, between resumes C and B

captures systemic discrimination—hiring decision disparities among equally-qualified workers perceived to be of the same group due to the non-group characteristics.

Outside of experimental settings, the core IA logic can be applied to the observable Y_i^0 case whenever direct discrimination can be reliably measured. For example, if manager signals S_i are directly observed by the econometrician then direct discrimination is directly identified, $\tau(s) = E[A_i \mid G_i = w, S_i = s] - E[A_i \mid G_i = b, S_i = s]$, and the direct discrimination components of equations (4)-(6) can be constructed from these and the conditional distribution of S_i given (G_i, Y_i^0). Subtracting one of these from the identified total discrimination measure again yields the corresponding measure of systemic discrimination.[37]

5.3 Selectively Observed or Proxied Y_i^0

In some cases, the researcher-chosen measure of qualification may be only selectively observed given the manager's actions. For example, when $Y_i^0 = Y_i^*$ measures a worker's productivity in the task at hand and $A_i \in \{0, 1\}$ indicates a hiring decision, observed output $Y_i = A_i Y_i^0$ gives a selective measure of qualification: workers who are hired ($A_i = 1$) reveal their qualification on the job but Y_i^0 is unobserved among unhired workers. Selective observability may also pose a challenge when Y_i^0 is an "upstream" measure of productivity, such as when a worker first enters the labor market prior to the hiring task at hand.

The IA approach to measuring systemic discrimination can be adopted to this case with additional (quasi-)experimental variation, appropriate to address the new selection challenge. This extension builds on Arnold et al. (2022), who develop quasi-experimental methods to study disparate impact in pretrial release decisions by leveraging the as-good-as-random assignment of individuals to bail judges. To translate their approach to the hiring example, suppose managers with potentially different hiring rates are as-good-as-randomly assigned to workers. Arnold et al. (2022) show how such assignment can be used to "selection-correct" the observed distribution of qualification by group, and how the resulting unselected qualification distribution can be used to estimate total discrimination by a particular adjustment of the unconditional group disparities $E[A_i \mid G_i = w] - E[A_i \mid G_i = b]$.[38] In practice, manager assignment can be substituted with any (quasi-)experimental variation in actions that allows for such correction of the selected observed qualification distribution.

To measure systemic discrimination in such settings, one can combine experimental variation in group membership perceptions—as in a classic audit or correspondence study—

[37] More generally, when S_i is only partially observed, variants of the frameworks of Altonji, Elder, and Taber (2005) and Oster (2019) may be applied to bound or point-identify direct discrimination from the change in disparities when only observed signals are conditioned on. Systemic discrimination is then also bounded or point-identified via equations (4)-(6).

[38] The Arnold et al. (2022) selection correction uses a non-parametric instrumental variables approach similar to Heckman (1990). While their method of estimating total discrimination uses the fact that Y_i^0 is binary, it can be extended to multivalued or continuous Y_i^0.

with the (quasi-)experimental action variation underlying the Arnold et al. (2022) approach. Specifically, consider a set of group-w workers with experimentally manipulated group perceptions among as-good-as-randomly assigned managers. The direct discrimination component in equation (4) could be estimated in this subsample by using the quasi-experimental selection correction technique to adjust the experimental disparities $E[A_i \mid \tilde{G}_i = b, G_i = w] - E[A_i \mid \tilde{G}_i = w, G_i = w]$. Subtracting this term from the Arnold et al. (2022) measure of total discrimination identifies the systemic discrimination component in equation (4). Analogous steps identify the other decompositions, as before.

A more challenging identification problem arises when Y_i^0 is not even selectively observed and must be proxied by other observables X_i. Here the frameworks of Altonji et al. (2005) and Oster (2019) may be integrated in the IA approach to measure systemic discrimination. Specifically, one can use these frameworks to bound or point-identify total discrimination from unconditional and conditional-on-X_i disparities by making assumptions about how the effect of conditioning with observables relates to the effect of the infeasible conditioning on Y_i^0. In samples where group membership is experimentally manipulated, such extrapolations may further bound or point-identify the direct discrimination component in equation (4) and therefore the corresponding systemic discrimination component. We leave the details of such extensions for future research.

To summarize, the IA approach can be used to bring each of the three decompositions (4)-(6) to different forms of data by leveraging different combinations of (quasi-)experimental and observational variation. We emphasize that the qualification metric Y_i^0 reflects the researcher's choice of the form of discrimination being studied, and different choices may require different sources of variation and identification strategies. We also note that with multiple choices of Y_i^0 it is possible to further decompose total discrimination into a direct component and multiple systemic components (perhaps reflecting different informational or technological sources). Bringing these richer decompositions to data would follow similarly as above, and likely require additional (quasi-)experimental variation.

6 Empirical Illustration

We now illustrate how our decomposition can be used to measure systemic discrimination and test predictions of our theory in two stylized lab experiments, conducted on the *Prolific.co* platform. The first experiment shows how systemic discrimination can arise from signal inflation, as in the motivating example in Section 2. The second experiment shows how systemic discrimination can be heterogeneous in screening decisions, similar to the example in Section 4.4.1. In both experiments, a pool of workers faced evaluations from two sets of managers. The first set of managers (termed Recruiters) generated initial evaluations of workers based on a productivity signal and their self-identified gender. The second set of managers (termed Hiring Managers) evaluated workers based on their self-identified gender and an endogenous productivity signal generated by the Recruiters. Worker qualification was chosen so that there is no systemic gender discrimination in Recruiter evaluations: total discrimination equals direct discrimination in this stage. Direct discrimination by Recruiters could lead to (informational) systemic discrimination in Hiring Manager' evaluations, alongside additional direct discrimination by Hiring Managers.

6.1 Signal Inflation

6.1.1 Setup

Workers: 100 participants were randomly assigned to the role of Worker. Each Worker completed two sets of tasks (A and B) and provided basic demographic information including self-reported group identity G_i (either male m or female f). Each task consisted of a test of the Worker's basic math, business, and history knowledge, with 10 randomly selected questions from these subjects. A Worker's performance on each task was defined as the number of questions she answered correctly. We restrict attention to Workers with a task-A performance in the range of 2-6 in order to ensure enough data for each gender.

Recruiters: 201 participants were randomly assigned to the role of Recruiter and given a budget of 10 experimental units (10 EU=$1 USD). Each Recruiter was shown information about two Workers and reported their highest willingness to pay to hire each. Specifically, Recruiters were shown the task-A performance of the Worker, which constituted their signal S_i^R, as well as the Worker's gender G_i. After viewing S_i^R and G_i, Recruiters were asked to state their willingness to pay to hire Worker i in the range of 0-10 EUs. This willingness to pay constituted the Recruiter action, A_i^R, with $\mathcal{A} = \{0, \ldots, 10\}$. Recruiter wage offers were then accepted or rejected according to the Becker-DeGroot-Marschak mechanism to incentivize truthful reporting: if a Worker was hired, Recruiters received 1 EU for each question the Worker answered correctly on task B minus the wage. If the Worker was not hired, the Recruiter did not pay anything and kept their endowment.[39]

Hiring Managers: 504 participants were randomly assigned to the role of Hiring Manager and also given a budget of 10 EU. Each Hiring Manager was shown the gender and Recruiter wage offer of a Worker.[40] Formally, each Hiring Manager observed signal $S_i^H \equiv A_{ik}^R$ for some Recruiter k assigned to Worker i, with $\mathcal{S}^H = \{0, \ldots, 10\}$. Hiring Managers then stated their maximum willingness to pay to hire the Worker using the same Becker-DeGroot-Marschak

[39]Here and in the next study we censor earnings at zero so that they could not be negative. Both Recruiters and Hiring Managers saw examples of the mechanism, examples of the task faced by the Workers, and passed comprehension checks before making wage offers.

[40]Hiring Managers saw only one Worker profile in order to minimize potential contrast effects.

Table 1. Signal Inflation: Recruiter and Hiring Manager Wage Offers

	Recruiter		Hiring Manager	
	(1)	(2)	(3)	(4)
Male Worker	0.49***	0.49***	0.94***	0.41**
	(0.12)	(0.12)	(0.21)	(0.18)
Productivity Signal		0.49***		0.56***
		(0.09)		(0.04)

Notes: This table reports coefficients from regressing Recruiter and Hiring Manager wage offers on Worker gender and a signal of Worker productivity. The productivity signal is the Worker's task-A performance in Column 2 and a Recruiter's wage offer to that Worker in Column 4. Columns 1 and 2 include 201 Recruiters, each evaluating two Workers. Columns 3 and 4 include 504 Hiring Managers, each evaluating one Worker. Standard errors, clustered at the manager level, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

mechanism as the Recruiters. We denote the Hiring Manager's action (wage offer) as A_i^H, with $\mathcal{A} = \{0, \ldots, 10\}$ as before.

6.1.2 Results

We measure discrimination with respect to a Worker's task-A performance Y_i^0, where $\mathcal{Y}^0 = \{2, 3, 4, 5, 6\}$. Setting Y_i^0 as task-A performance focuses attention on disparities among Workers who enter the hiring market with the same initial observable signals of productivity.

Workers: There were no significant gender differences in Worker performance on either task. On average, Workers completed 3.57 questions correctly on task A and 3.53 questions correctly on task B. Regressing overall performance (the sum of performance on both tasks) on a male Worker indicator yields an insignificant coefficient of -0.13 ($p = 0.84$). The gender coefficient is similarly insignificant when we regress performance on task A (0.21; $p = 0.63$) and task B (-0.34; $p = 0.35$) separately. Performance on task B was predictive of performance on task A. Regressing the latter on the former yields a coefficient of 0.36 ($p < 0.01$). Furthermore, there were no significant gender differences in this relationship: regressing task-A performance on task-B performance, gender, and their interaction yields an insignificant interaction coefficient of 0.15 ($p = 0.58$).

Recruiters: Since Worker qualification Y_i^0 coincides with the Recruiter signal S_i^R, any discrimination in the initial evaluations is direct. We can rule out accurate statistical discrimination as a driver of such direct discrimination, as the signal is equally informative of Worker productivity for both men and women. Any direct discrimination by Recruiters is therefore driven by biased preferences or beliefs.

Recruiters directly discriminated against female Workers. The average offered wage was 5.23. Column 1 of Table 1 shows that male Workers were on average offered a 0.49 higher wage than female Workers ($p < 0.01$).[41] This effect corresponds to around 0.22 standard deviations of Recruiter wage offers. Column 2 shows that Recruiters responded positively to their signal, with each additional question correctly answered in task A leading to a higher wage offer of 0.49 on average ($p < 0.01$).[42] While this data alone cannot be used to disentangle preference and belief-based sources of direct discrimination, it is consistent with prior work showing inaccurate beliefs or stereotypes as drivers of gender discrimination in similar settings (Bordalo et al. 2019; Bohren et al. 2019).

Hiring Managers: Since G_i is independent of Y_i^0, any disparities in Hiring Manager wage offers A_i^H reflect discrimination. Such discrimination could be direct (i.e., among male and female Workers with the same Hiring Manager signal realization S_i^H) or systemic (i.e., stemming from male and female Workers with the same Recruiter signal realization who then receive different Recruiter wage offers).

Hiring Managers discriminated against female Workers. The average Hiring Manager wage offer was 5.50. Column 3 of Table 1 shows that male Workers were on average offered a 0.94 higher wage than female Workers ($p < 0.01$). This disparity captures corresponds to roughly 0.39 standard deviations of Hiring Manager wage offers.

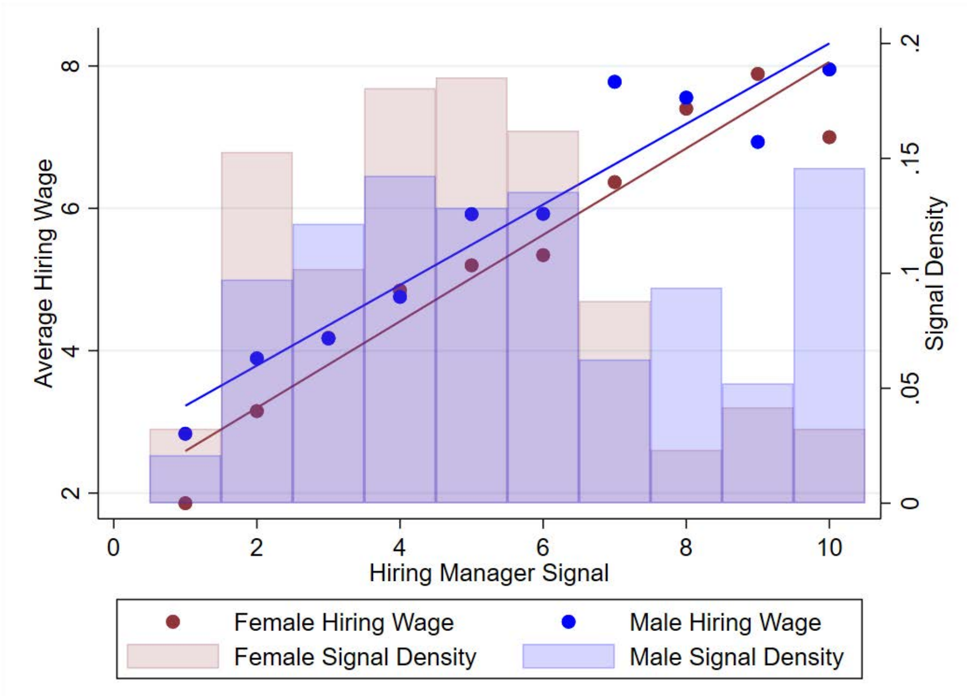
Column 4 of Table 1 further suggests that much of the discrimination by Hiring Managers is systemic. Controlling for the Hiring Manager signal (i.e., the Recruiter wage offer) decreases the gender coefficient to 0.41 ($p = 0.02$). Interestingly, both the gender coefficient and productivity signal coefficient are similar to those in Column 2.

Figure 2 illustrates the two sources of Hiring Manager discrimination. The scatter plot shows average Hiring Manager wages as a function of the Worker's gender and productivity signal. The lines of best fit show a positive relationship between the signal and wage for both genders. This relationship is shifted upward for male Workers, illustrating direct discrimination: conditional on seeing the same signal, a male Worker received a higher wage than a female Worker. Importantly, however, the *distribution* of productivity signals differs by gender: male Workers tend to have higher signals than female Workers, due to direct discrimination in initial evaluations. Since higher signals lead to higher wages from the Hiring Managers (the upward sloping lines), this pattern leads to systemic discrimination.

We now quantify systemic discrimination using the decompositions in Section 5.1. We first estimate total Hiring Manager discrimination $\Delta(y^0)$ by comparing male and female wage offers for each task-A performance level. We then estimate Hiring Managers' average direct discrimination with a given task-A performance, $E[\tau_i \mid G_i = w, Y_i^0 = y^0]$, by averaging

[41] Since Recruiters made offers to multiple Workers, standard errors are clustered at the Worker level.

[42] The coefficient without the gender control is identical, 0.49 ($p < 0.01$), since G_i and S_i^R are uncorrelated.



Notes: This figure plots average Hiring Manager wage offers for female and male signals with different productivity signals (on the left y-axis) and the distribution of female and male productivity signals (on the right y-axis). Gender differences in the former illustrate direct discrimination, while gender differences in the latter illustrate the source of systemic discrimination.

FIGURE 2. Signal Inflation: Hiring Manager Wage Offers by Worker Gender and Signal

gender disparities across each Hiring Manager signal realization according to the distribution each task-A performance induces over the Hiring Manager signal (i.e., the Recruiter wage offer). Per Equation (4), subtracting this estimate of direct discrimination from the estimate of total discrimination yields an estimate of systemic discrimination at each qualification level y. We then average these measures of total, direct, and systemic discrimination over the marginal distribution of Worker qualification by gender before equal-weighting these gender-specific averages.[43] We similarly decompose total discrimination into the alternative measures of direct and systemic components in Equations (5) and (6). See Appendix A for details on these three calculations.

Table 2 confirms significant systemic discrimination in Hiring Manager wage offers. Estimated total discrimination against female Workers averages to 0.90, similar to the regression estimate in Column 3 of Table 1. Estimated average direct discrimination is around 0.41 for each decomposition, similar to the regression estimate in Column 4 of Table 1. Estimated systemic discrimination is around 0.49 for each of the three decompositions. The majority (54%) of discrimination thus comes from signal inflation, as with the model in Section 2.

[43]Results are similar for other weighting schemes, such as by the overall qualification distribution.

Table 2. Signal Inflation: Total, Direct, and Systemic Discrimination in Manager Wages

	(1)	(2)	(3)
Total Discrimination	0.90***	0.90***	0.90***
	(0.22)	(0.22)	(0.22)
Average Direct Discrimination	0.41*	0.41**	0.41**
	(0.23)	(0.18)	(0.19)
Systemic Discrimination	0.49**	0.49***	0.49***
	(0.22)	(0.16)	(0.18)
Decomposition Method	Equation (4)	Equation (5)	Equation (6)

Notes: This table reports estimates of total discrimination, average direct discrimination, and systemic discrimination in Hiring Manager wage offers. Total discrimination is measured by the average difference in offers among male vs. female Workers with a given task-A score, averaged by the equal-weighted distribution of task-A scores for male and female Workers. Average direct and systemic discrimination are measured by equations Equations (4) to (6), then similarly averaged by the distribution of task-A scores. The sample includes 504 Hiring Managers, each evaluating one Worker. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

6.2 Screening

The same group of Workers were retained for the second experiment. A new group of 199 Recruiters were shown the task-A performance of two Workers, along with the Workers' gender, and asked to select which Worker they would prefer to hire. Recruiters were then paid 1 USD for each question the hired Worker answered correctly on task B, above 5. The Recruiter's action rule is thus $A_i^R \in \{0, 1\}$.

A new group of 501 Hiring Managers saw one Worker's profile after their evaluation by a Recruiter, along with the Worker's gender. They were shown information on the Worker's task-A performance only if the Recruiter had chosen to hire them; otherwise they saw no performance information. Hiring Managers then made a binary decision of whether or not to hire the Worker. If the Worker was hired, the Hiring Manager received a bonus corresponding to their task-B performance; otherwise, the Hiring Manager received 4 dollars with certainty.

Formally, each Hiring Manager j observed a signal S_i^H corresponding to Worker i's task-A performance if the Worker was hired by the recruiter ($A_i^R = 1$). If the Worker was not hired ($A_i^R = 0$), the Hiring Manager observed no signal ($S^H = \emptyset$). Recruiter actions thus affected the *informativeness* of Hiring Manager signals—whether or not she saw an objective signal of productivity. This setting was thus designed to emulate the process by which managers can obtain more accurate performance signals depending on whether potential Workers had access to prior opportunities to "prove themselves" (e.g., internships). The Manager's action $A_i^H \in \{0, 1\}$ corresponds to her hiring the Worker.

6.2.1 Results

As before, we measure systemic and total discrimination with respect to task-A performance Y_i^0; since this qualification measure again coincides with the Recruiter signal, any discrimination in the initial hiring stage is direct. Discrimination in the Hiring Manager stage can again be direct or systemic. Per Section 4.4.1, we expect the differences in signal informativeness to lead to heterogeneity in systemic discrimination by Workers qualification.

Recruiters: Recruiters directly discriminated against female Workers. The hiring rate for male Workers was 28 percentage points higher than for female Workers ($p < 0.01$), who were hired at a rate of 36%.[44] Given the lack of gender-based performance differences, as reported in Section 6.1.2, this disparity in hiring rates is not consistent with accurate statistical discrimination. Therefore, Recruiter direct discrimination again stems from either biased preferences or beliefs.

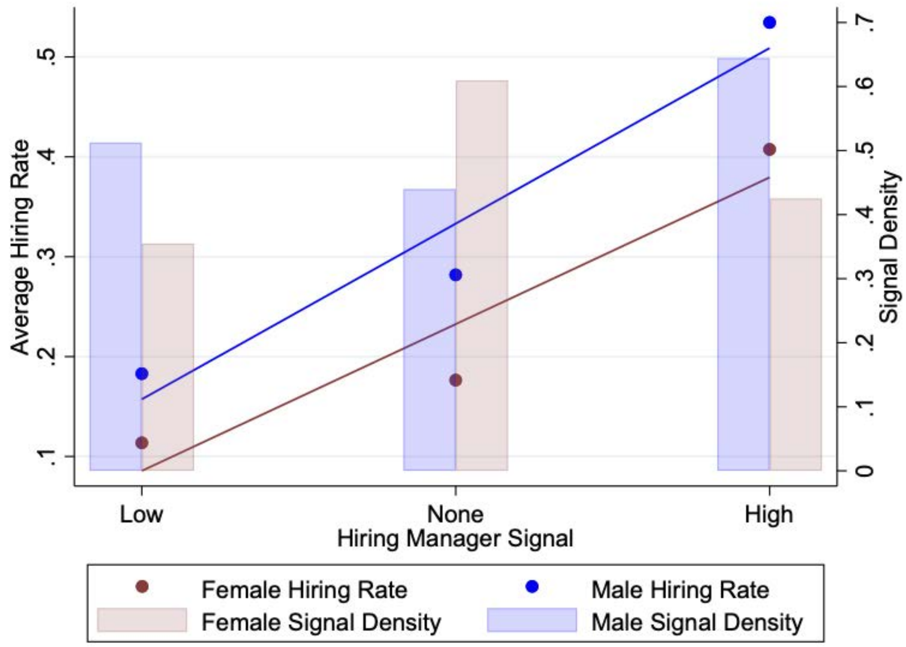
Hiring Managers: Hiring Managers discriminated against female Workers. On average, male Workers were hired at a 9 percentage point higher rate than female Workers ($p = 0.02$), who were hired at a rate of 0.22. However, this average effect masks important heterogeneity. Among Workers with low (below-median) qualification levels, male Workers were hired at an insignificant 4 percentage point higher rate ($p = 0.43$).[45] Among Workers with high (above-median) qualification levels, male Workers were hired at a significant 23 percentage point higher rate ($p < 0.01$).

Figure 3 illustrates the reason for this heterogeneity in total discrimination. Similar to Figure 2, the scatter plot shows the average Hiring Manager actions conditional on the signal (or lack thereof) and the Worker's gender. As before, the lines of best fit show a positive relationship between the signal and the probability of getting hired for both groups: Hiring Managers were more likely to hire a Worker after seeing a high signal than a low signal, with the hiring rate for no signal laying in between. Conditional hiring rates are shifted upward for male Workers, illustrating direct discrimination. Importantly, however, the distribution of signals seen by Managers also differs by gender: direct discrimination by Recruiters made Managers more likely to see both low and high signals from male Workers than female Workers, with female Workers being much more likely to have an uninformative signal. Given the upward-sloping lines, female Workers with high qualification levels were likely to be hurt by systemic discrimination, while female Workers with low qualification levels were likely to be helped by it.

As before, we quantify total, direct, and systemic discrimination in Hiring Manager actions using the decompositions in Section 5.1. We first estimate total Hiring Manager dis-

[44] Standard errors are clustered at the individual level.

[45] The median task-A performance was 4.



Notes: This figure plots average Hiring Manager wage offers for female and male signals with different productivity signals (on the left y-axis) and the distribution of female and male productivity signals, conditional on qualification (on the right y-axis). Gender differences in the former illustrate direct discrimination, while gender differences in the latter illustrate the source of systemic discrimination.

FIGURE 3. Screening: Hiring Manager Wage Offers by Worker Gender and Signal

crimination $\Delta(y^0)$ by comparing male and female hiring rates based on task-A performance. We then estimate the average direct Hiring Manager discrimination $E[\tau_i \mid G_i = w, Y_i^0 = y^0]$ faced by male Workers with a given task-A performance by averaging gender disparities across each Hiring Manager signal realization according to the distribution each task-A performance induces over this signal. Subtracting this estimate of from the estimate of total discrimination yields an estimate of the measure of systemic discrimination.[46] We average these measures over the distribution of task-A performance as before, separately for Workers with low (below-median) and high (above-median) qualification levels.

Table 3 confirms the heterogeneity in systemic discrimination faced by women with different qualification levels. For highly qualified women, total discrimination is estimated as a significant 0.24. Our decomposition shows this is driven by a combination of significant direct (0.15) and systemic discrimination (0.09). In contrast, total discrimination among less qualified Workers is small and insignificant (0.03), despite significant direct discrimination. The reason is a small degree of negative systemic discrimination among less qualified Workers (-0.04). Consistent with the model in Section 4.4.1, the gap in systemic discrimination across qualification levels is significant ($p = 0.04$).

[46]Here we use the "average" decomposition, Equation (5). The other decompositions give similar results.

Table 3. Screening: Total, Direct, and Systemic Discrimination in Hiring Manager Actions

	High Qualification (1)	Low Qualification (2)	Difference (3)
Total Discrimination	0.24***	0.03	0.21***
	(0.06)	(0.04)	(0.07)
Average Direct Discrimination	0.15***	0.07**	0.08
	(0.05)	(0.04)	(0.05)
Systemic Discrimination	0.09**	-0.04	0.13**
	(0.04)	(0.03)	(0.06)

Notes: This table reports estimates of total discrimination, average direct discrimination, and systemic discrimination in Hiring Manager hiring rates across different levels of Worker performance on Task A. Total discrimination is measured by the average difference in hiring rates among male vs. female Workers with a given task-A score, averaged by the equal-weighted distribution of task-A scores for male and female Workers. Average direct and systemic discrimination are measured by Equation (6), then similarly averaged by the distribution of task-A scores. The sample includes 501 Hiring Managers, each evaluating one Worker. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

In summary, our two lab experiments illustrate both the potential impact of systemic factors in treatment disparities (despite no underlying disparity in Worker productivity) as well as how such systemic discrimination can be measured. Importantly, despite the substantial levels of total discrimination in our setting, standard tools such as correspondence and audit studies would not have detected the majority of discrimination in Hiring Manager wage offers or hiring rates: direct Hiring Manager discrimination, which conditions on the non-gender signal, was smaller than total discrimination in the first study and missed important heterogeneity in total discrimination in the second study. The results also underscore the pitfalls of conditioning on observables that may themselves be the outcomes of previous discrimination; this strategy would suggest minimal discrimination in the first study, despite substantial total discrimination. Finally, the studies illustrate how direct discrimination against members of specific groups, such as those stemming from animus, inaccurate stereotypes, or accurate statistical discrimination (Becker 1957; Phelps 1972; Bordalo et al. 2016), can perpetuate total discrimination even when the direct discrimination is mitigated (as in Section 2 and Appendix B.1). Therefore policies which aim to eliminate direct discrimination through contact (Rao 2019; Paluck, Green, and Green 2019) or correcting beliefs (Bohren et al. 2022) may still allow discrimination to persist through systemic factors.

7 Iterated Audit

Our lab-in-the-field experiment follows the IA design developed in Section 5. We used a hiring and recruitment agency to recruit hiring managers ($N = 208$) with experience in

(A) No Work Experience

EDUCATION
Waterford High School
2013 - 2017

WORK EXPERIENCE
Pet Sitter
2018 – Present
- Provide pet sitting services including dog walking, feeding and yard care

VOLUNTEER EXPERIENCE
Community Soup kitchen

SKILLS
Proficient with Microsoft Word, Excel, and PowerPoint

(B) Work Experience

EDUCATION
Canyon View High School
2015 - 2019

WORK EXPERIENCE
Warehouse worker
2021 - Present
- Responsible for receiving, storing, and distributing goods.
- Loaded and unloaded trailers, order picked coordinated material transfers, and replenish slots that were low in materials.

Pet Sitter
2018 – Present
- Provide pet sitting services including dog walking, feeding and yard care

VOLUNTEER EXPERIENCE
Community Food bank

SKILLS
Proficient with Microsoft Word, Excel, and PowerPoint

Notes: This figure shows example resumes in the iterated audit. Panel A shows a resume with no relevant work experience, while Panel B shows the same resume with added work experience.

FIGURE 4. Iterated Audit: Example Resumes

evaluating applicants to entry-level jobs and who were currently looking for employees.[47] In a factorial ratings design based on Kessler et al. (2019), the hiring managers evaluated fictitious resumes to an entry level job on the likelihood of the applicant being hired for the job.[48] This 1-10 hiring likelihood score is the main dependent variable in our analyses. Decisions were incentivized using a similar methodology as Kessler et al. (2019). Although the resumes themselves were fictitious, the components (e.g., prior work experience) could be matched to resumes of actual potential applicants to the relevant job. In this way, likelihood scores were incentivized because a hiring manager's highest-rated fictitious applicant could be matched to an actual applicant with similar qualifications.

Unlike a standard correspondence or audit study that presented evaluators with two sets of resumes—identical versions that only differed on signals of group identity—our IA design featured three sets of resumes drawn from groups A, B, C as depicted in Figure 1. Two of the three sets (A and C) differed on group identity—featuring distinctively Black or white names—as well as previous work experience. Example resumes are depicted in Figure 4, where Figure 4b features relevant work experience while Figure 4a does not. The set of resumes B had the same level of work experience as set A but differed on group identity.

The proportion of resumes with previous work experience was constructed based on the results of a previously run audit study by Pager (2003). There, matched pairs of individuals applied for entry level jobs. Black applicants were found to be significantly less likely to proceed through the application process than white applicants with the same qualifica-

[47]Hiring managers had an average of 6.7 years of experience in their current role.

[48]See Lahey and Oxley (2021) and Kübler et al. (2018) for similar uses of factorial resume designs in studying discrimination.

tions. We use this finding difference to generate the proportion of white-endogenous (set A) and Black-endogenous (set C) resumes that have previous work experience in our study. Specifically, 32% of white-endogenous resumes had relevant previous work experience compared to 18% of Black-endogenous resumes. The third set of resumes (B) are identical to the white-endogenous (A) set but feature distinctively Black names; we refer to this set as Black-exogenous. We generated six different resumes for each set A, B, and C.

Each hiring manager evaluated a random draw of 4 different resumes. The managers were told about the incentive structure. Importantly, they were also informed about the results of the Pager (2003) study, particularly the differential rates with which white versus Black applicants were contacted to proceed with the entry level job. Managers were thus informed about potential racial disparities in signals, i.e., previous work experience in entry-level jobs. This design choice likely yields a conservative test for systemic discrimination in this setting.

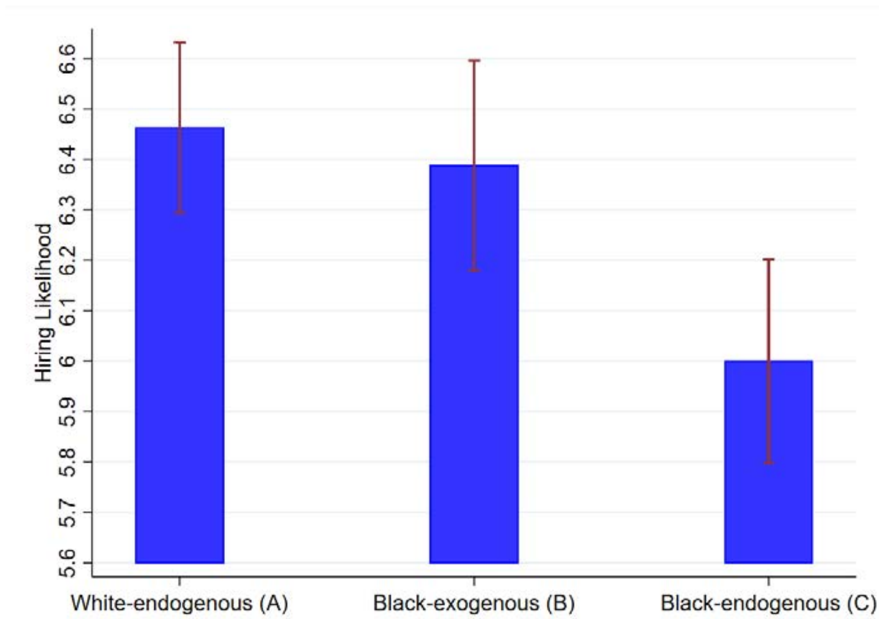
7.1 Results

Comparing evaluations of the two endogenous sets of resumes (A vs. C) identifies total discrimination, while comparing white-endogenous resumes to Black-exogenous resumes (A vs. B), the standard comparison in correspondence studies, gives a measure of direct discrimination. Our decomposition then allows us to measure systemic discrimination.

Figure 5 presents results on the average hiring likelihoods by resume type. Table 4 presents these results in regression form. We find significant total discrimination. Column 1 of the table compares white-endogenous (A) to Black-endogenous (C) resumes and reveals a gap of 0.46 in the likelihood of getting hired. This corresponds to 21% of one standard deviation in hiring likelihood. Similar to the study on signal inflation, we can examine how much of the total discrimination is driven by non-group characteristics by including a control for prior work experience. Column 2 shows that prior work experience has a large and significant impact on hiring likelihood. Moreover, consistent with total discrimination being driven in large part by race-based differences in signals (systemic discrimination), the gap in hiring likelihood decreases substantially and becomes insignificant when controls for prior experience are added.

Comparing white-endogenous (A) to Black-exogenous (B) in Column 3 reveals that the majority of total discrimination is driven by systemic factors. While the coefficient on race is negative, it is small and not statistically significant. The majority of total discrimination is driven by race-based differences in prior work experience.

Strikingly, the systemic discrimination in prior work experience impacts behavior *despite* the hiring managers being told that they were likely generated by direct discrimination elsewhere in the system. While this information likely reduced the extent of direct discrimination (per Column 3 of Table 4), systemic discrimination still led to substantial differences in hiring likelihoods. In light of prior work showing the effectiveness of information in re-



Notes: This figure plots average hiring likelihoods for the three resume types in the iterated audit experiment. The gap between white-endogenous and Black-endogenous hiring rates illustrates total discrimination. The gap between white-endogenous and Black-exogenous hiring rates illustrates average direct discrimination. The gap between Black-exogenous and Black-endogenous hiring rates illustrates systemic discrimination. Whiskers indicate +/- one manager-clustered standard error.

FIGURE 5. Iterated Audit: Total, Direct, and Systemic Discrimination in Hiring Likelihoods

ducing direct discrimination (Bohren et al. 2022), these findings highlight the difficulty of mitigating total discrimination when it is caused by systemic factors.

8 Conclusion

Vast literatures, mostly from outside of economics, emphasize the importance of systemic factors in driving group-based disparities, yet economic analyses largely focus on direct discrimination on the basis of group identity itself. This paper begins to bridge this gap by developing new theoretical and empirical tools to study systemic discrimination. We show how economic models and measures of individual direct discrimination can be seen as focusing on one component of total discrimination. This analysis suggests high returns to new economic theories of how systemic discrimination can arise and persist across different contexts and time periods. Our decomposition of total discrimination into direct and systemic components further motivates the development of new econometric tools that identify these components with different forms of experimental and observational data. Our hiring experiments and novel Iterated Audit design show how conventional methods of studying direct discrimination can miss total discrimination and important heterogeneity in practice.

Table 4. Iterated Audit: Hiring Likelihoods

	(1)	(2)	(3)
Black Worker	-0.46**	-0.29	-0.08
	(0.20)	(0.20)	(0.21)
Prior Experience		1.26***	
		(0.21)	
Exogenous-Black Resumes	N	N	Y

Notes: This table reports coefficients from regressing hiring likelihoods on resume characteristics in the iterated audit. Columns 1 and 2 compare endogenous-Black and endogenous-white resumes (types A and C in Figure 1). Column 3 compares exogenous-Black and endogenous-White resumes (types A and B in Figure 1). The sample includes 208 Hiring Managers, each evaluating four different resumes. Manager-clustered standard errors are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

Understanding the interaction between different sources of direct and systemic discrimination is important from a policy perspective. As an example, consider the case of Ban-the-Box (BTB) policies that seek to eliminate questions about prior criminal history from job applications. The premise is based on the fact that employers are less likely to call back and interview applicants with past criminal records, even when those infractions are minor and not relevant for the job Stoll (2009); Finlay (2008). As we formalize, direct discrimination in policing will thus generate systemic discrimination from signal inflation against Black workers. BTB policies presumably address this disparity by eliminating the inflated signal. However, if evaluators believe that Black workers have a higher underlying propensity for criminal activity than white workers—then this can interact with screening discrimination to exacerbate disparities. Specifically, by making the applicant's signal less informative, BTB policies may lead employers to rely on their biased priors—hurting Black applicants without criminal records without necessarily helping those with criminal records. Agan and Starr (2017a) report results from a field experiment demonstrated this effect: removing information about criminal records exacerbated the Black-white callback gap from 7% to 43%. Doleac and Hansen (2020) report a similar effect using observational data. By formalizing the interaction between direct and systemic sources of discrimination, our framework is useful for interpreting and predicting the effects of policies that aim to address it.

New analytic tools may broaden the scope for formulating appropriate policy responses to the many large and persistent disparities documented in the literature. Indirect discrimination can lead to illegal disparate impact in some settings, as in the landmark *Griggs v. Duke Power Co. (1970)* finding. The development of robust econometric methods for measuring systemic and total discrimination, perhaps across different qualification measures, can be a powerful complement to existing regulatory tools in such settings.[49] Robust economic models of systemic discrimination can aide the interpretation of these methods, by enriching policymakers' understanding of dynamics and heterogeneity within and across different domains. Such theoretical and empirical advancements can improve policy making and equity in labor markets, housing, criminal justice, education, healthcare, and other areas.

[49]For example, the U.S. Equal Employment Opportunity Commission (EEOC) launched nearly 600 investigations into systemic discrimination in 2020. Many employment practices EEOC flags for possible systemic are indirect (such as word-of-mouth recruitment practices), and would thus not be picked up by a conventional correspondence or audit study (see https://www.eeoc.gov/systemic-enforcement-eeoc).

References

Aaronson, D., D. Hartley, and B. Mazumder (2021): "The Effects of the 1930s HOLC 'Redlining' Maps," *American Economic Journal: Economic Policy*, 13, 355–92.

Agan, A. and S. Starr (2017a): "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment," *The Quarterly Journal of Economics*, 133, 191–235.

——— (2017b): "The Effect of Criminal Records on Access to Employment," *American Economic Review P&P*, 107, 560–64.

Aigner, D. and G. Cain (1977): "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30, 175–187.

Allard, S. W. and M. L. Small (2013): "Reconsidering the Urban Disadvantaged: The Role of Systems, Institutions, and Organizations," *Annals of the American Academy of Political and Social Science*, 647, 6–20.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of political economy*, 113, 151–184.

Andre, P. (2022): "Shallow Meritocracy," *Working Paper*.

Angwin, J., J. Larson, S. Mattu, and L. Kirchner (2016): "Machine Bias," *ProPublica Report*.

Antonovics, K. and B. G. Knight (2009): "A New Look at Racial Profiling: Evidence from the Boston Police Department," *The Review of Economics and Statistics*, 91, 163–177.

Arnold, D., W. Dobbie, and P. Hull (2021): "Measuring Racial Discrimination in Algorithms," *AEA Papers and Proceedings*, 111, 49–54.

——— (2022): "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial Bias in Bail Decisions," *Quarterly Journal of Economics*, 133, 1885–1932.

Arrow, K. J. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter and A. Rees, Princeton, NJ: Princeton University Press.

Avery, M., A. Leibbrandt, and J. Vecci (2023): "Doe Artificial Intelligence Help or Hurt Gender Diversity? Evidence from Two Field Experiments on Recruitment in Tech," *Working Paper*.

Barron, K., R. Ditlmann, S. Gehrig, and S. Schweighofer-Kodritsch (2020): "Explicit and Implicit Belief-Based Gender Discrimination: a Hiring Experiment," *WZB Discussion Paper*.

Bartik, A. W. and S. T. Nelson (2016): *Credit reports as resumes: The incidence of pre-employment credit screening*, Massachusetts Institute of Technology, Department of

Economics.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 106, 1437–75.

Bayer, P., F. Ferreira, and S. L. Ross (2017): "What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders," *Review of Financial Studies*, 31, 175–205.

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

Benson, A., S. Board, and M. Meyer-ter Vehn (2019): "Discrimination in Hiring: Evidence from Retail Sales," *UCLA Working Paper*.

Berk, R., H. Heidari, S. Jabbari, M. Kearns, and A. Roth (2018): "Fairness in Criminal Justice Risk Assessments: The State of the Art," *Sociological Methods & Research*, 50, 1–42.

Bertrand, M. and E. Duflo (2016): *Field Experiments on Discrimination*, Elsevier.

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, 991–1013.

Blau, F. D. and L. M. Kahn (2017): "The Gender Wage Gap: Extend, Trends, and Explanations," *Journal of Economic Literature*, 55, 789–865.

Blinder, A. S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8, 436–455.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2022): "Inaccurate Statistical Discrimination: An Identification Problem," *Review of Economics and Statistics*.

Bohren, J. A. and D. N. Hauser (2022): "Representing Biases and Heuristics as Misspecified Models," *Working Paper*.

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

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016): "Stereotypes," *The Quarterly Journal of Economics*, 131, 1753–1794.

——— (2019): "Beliefs About Gender," *American Economic Review*, 109, 739–73.

Brekoulakis, S. (2013): "Systemic Bias and the Institution of International Arbitration: A New Approach to Arbitral Decision-Making," *Journal of International Dispute Settlement*, 4, 553–585.

Brownback, A. and M. A. Kuhn (2019): "Understanding Outcome Bias," *Games and Economic Behavior*, 117, 342–360.

Buolamwini, J. (2022): "Facing the Coded Gaze with Evocative Audits and Algorithmic Audits," Ph.D. thesis, Massachusetts Institute of Technology.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017): "'Acting Wife': Marriage Market Incentives and Labor Market Incentives," *American Economic Review*, 107, 3288–3319.

Canay, I., M. Mogstad, and J. Mountjoy (2020): "On the Use of Outcome Tests for Detecting Bias in Decision Making," *NBER Working Paper No. 27802*.

Card, D. and A. B. Krueger (1992): "School Quality and Black-White Relative Earnings: A Direct Assessment," *Quarterly Journal of Economics*, 107, 151–200.

Chandra, A. and D. O. Staiger (2010): "Identifying Provider Prejudice in Healthcare," *NBER Working Paper No. 16382*.

Charles, K. K. and E. Hurst (2002): "The Transition To Home Ownership And The Black-White Wealth Gap," *Review of Economics and Statistics*, 84, 281–297.

Coate, S. and G. C. Loury (1993): "Will Affirmative-Action Polices Eliminate Negative Stereotypes?" *American Economic Review*, 83, 1220–1240.

Coffman, K. B., C. L. Exley, and M. Niederle (2021): "The Role of Beliefs in Driving Gender Discrimination," *Management Science*, 67, 3551–3569.

Cook, L. (2014): "Violence and Economic Growth: Evidence from African American Patents, 1870-1940," *Journal of Economic Growth*, 19, 221–257.

Darity, W. (2005): "Stratification Economics: the Role of Intergroup Inequality," *Journal of Economics and Finance*, 29, 144–153.

Darity, W. A. and P. L. Mason (1998): "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender," *Journal of Economic Perspectives*, 12, 63–90.

De Plevitz, L. (2007): "Systemic Racism: The Hidden Barrier to Educational Success for Indigenous School Students," *Australian Journal of Education*, 51, 54–71.

De Quidt, J., J. Haushofer, and C. Roth (2018): "Measuring and Bounding Experimenter Demand," *American Economic Review*, 108, 3266–3302.

Doleac, J. L. and B. Hansen (2020): "The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden," *Journal of Labor Economics*, 38, 321–374.

Fang, H. and A. Moro (2011): "Theories of Statistical Discrimination and Affirmative Action: A Survey," in *Handbook of Social Economics*, Elsevier, vol. 1, 133–200.

Farkas, G. (2003): "Racial Disparities and Discrimination in Education: What Do We Know, How Do We Know It, and What Do We Need to Know?" *Teachers College Record*, 105, 1119–1146.

Feagin, J. (2013): *Systemic Racism: A Theory of Oppression*, Routledge.

Feagin, J. R. and C. B. Feagin (1978): *Discrimination American Style: Institutional Racism and Sexism*, Prentice Hall.

Finlay, K. (2008): "Effect of employer access to criminal history data on the labor market outcomes of ex-offenders and non-offenders," Tech. rep., National Bureau of Economic

Research.

Fiske, S. T. (1998): "Stereotyping, Prejudice, and Discrimination," *The Handbook of Social Psychology*, 2, 357–411.

Fisman, R., D. Paravisini, and V. Vig (2017): "Cultural proximity and loan outcomes," *American Economic Review*, 107, 457–92.

Fryer, R. G. and S. D. Levitt (2004): "The Causes and Consequences of Distinctively Black Names," *The Quarterly Journal of Economics*, 119, 767–805.

Gaddis, S. M. (2017): "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies," *Sociological Science*, 4, 469–489.

Gallen, Y. and M. Wasserman (2021): "Informed Choices: Gender Gaps in Career Advice," *CEPR Discussion Paper No. DP15728*.

Gebru, T. (2020): "Race and gender," *The Oxford handbook of ethics of aI*, 251–269.

Glover, D., A. Pallais, and W. Pariente (2017a): "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores," *The Quarterly Journal of Economics*, 132, 1219–1260.

——— (2017b): "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores," *The Quarterly Journal of Economics*, 132, 1219–1260.

Gorman, E. H. (2005): "Gender Stereotypes, Same-Gender Preferences, and Organizational Variation in the Hiring of Women: Evidence from Law Firms," *American Sociological Review*, 70, 702–728.

Grau, N. and D. Vergara (2021): "An Observational Implementation of the Outcome Test with an Application to Ethnic Prejudice in Pretrial Detentions," *Working Paper*.

Gynter, P. (2003): "On the Doctrine of Systemic Discrimination and its Usability in the Field of Education," *International Journal on Minority and Group Rights*, 10, 45–54.

Hardt, M., E. Price, and N. Srebro (2016): "Equality of Opportunity in Supervised Learning," *Proceedings of the 30th Conference on Neural Information Processing Systems*, 3323–3331.

Heckman, J. J. (1990): "Varieties of Selection Bias," *American Economic Review Papers and Proceedings*, 80, 313–318.

Hill, R. B. (1988): *Structural Discrimination: The Unintended Consequences of Institutional Processes.*, Wesleyan University Press.

Hübert, R. and A. T. Little (2020): "A Behavioral Theory of Discrimination in Policing," *Working Paper*.

Hull, P. (2021): "What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making," *NBER Working Paper No. 28503*.

Hurst, E., Y. Rubinstein, and K. Shimizu (2021): "Task-based discrimination," Tech. rep., National Bureau of Economic Research.

Jolly-Ryan, J. (1998): "Chipping Away at Discrimination at the Country Club," *Pepperdine Law Review*, 25, 2.

Kasy, M. and R. Abebe (2021): "Fairness, equality, and power in algorithmic decision-making," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 576–586.

Kessler, J. B., C. Low, and C. D. Sullivan (2019): "Incentivized resume rating: Eliciting employer preferences without deception," *American Economic Review*, 109, 3713–44.

Kitagawa, E. M. (1955): "Components of a Difference Between Two Rates," *Journal of the American Statistical Association*, 50, 1168–1194.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018): "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 133, 237–293.

Kline, P. M., E. K. Rose, and C. R. Walters (2021): "Systemic Discrimination among Large U.S. Employers," *NBER Working Paper No. 29053*.

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

Kohler-Hausmann, I. (2019): "Eddie Murphy and the Dangers of Counterfactual Causal Thinking about Detecting Racial Discrimination," *Northwestern University Law Review*, 113, 1163–1227.

Kübler, D., J. Schmid, and R. Stüber (2018): "Gender discrimination in hiring across occupations: a nationally-representative vignette study," *Labour Economics*, 55, 215–229.

Lahey, J. N. and D. R. Oxley (2021): "Discrimination at the Intersection of Age, Race, and Gender: Evidence from an Eye-Tracking Experiment," *Journal of Policy Analysis and Management*, 40, 1083–1119.

List, J. A. (2004): "The nature and extent of discrimination in the marketplace: Evidence from the field," *The Quarterly Journal of Economics*, 119, 49–89.

Mayhew, L. H. (1968): *Law and Equal Opportunity*, Harvard University Press.

Mengel, F., J. Sauermann, and U. Zölitz (2019): "Gender Bias in Teaching Evaluations," *Journal of the European Economic Association*, 17, 535–566.

Miller, D. T. and C. McFarland (1987): "Pluralistic Ignorance: When Similarity is Interpreted as Dissimilarity." *Journal of Personality and Social Psychology*, 53, 298.

Mocanu, T. (2022): "Designing Gender Equity: Evidence from Hiring Practices and Committees," .

Mustard, D. B. (2001): "Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the US Federal Courts," *The Journal of Law and Economics*, 44, 285–314.

Nazroo, J. Y. (2003): "The Structuring of Ethnic Inequalities in Health: Economic Position, Racial Discrimination, and Racism," *American Journal of Public Health*, 93, 277–284.

Neal, D. A. and W. R. Johnson (1996): "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104, 869–895.

Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709.

Oster, E. (2019): "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 37, 187–204.

Pager, D. (2003): "The mark of a criminal record," *American journal of sociology*, 108, 937–975.

Pager, D., B. Bonikowski, and B. Western (2009): "Discrimination in a Low-Wage Labor Market: A Field Experiment," *American Sociological Review*, 74, 777–799.

Pager, D. and H. Shepherd (2008): "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets," *Annuual Review of Sociology*, 34, 181–209.

Paluck, E. L., S. A. Green, and D. P. Green (2019): "The Contact Hypothesis Re-evaluated," *Behavioural Public Policy*, 3, 129–158.

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

Pierson, E., C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, et al. (2020): "A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States," *Nature Human Behaviour*, 4, 736–745.

Pincus, F. L. (1996): "Discrimination Comes in Many Forms: Individual, Institutional, and Structural," *American Behavioral Scientist*, 40, 186–194.

Powell, J. A. (2007): "Structural Racism: Building Upon the Insights of John Calmore," *North Carolina Law Review*, 86, 791.

Rambachan, A. and J. Roth (2020): "Bias In, Bias Out? Evaluating the Folk Wisdom," *1st Symposium on the Foundations of Responsible Computing (FORC 2020)*, 156, 6:1–6:15.

Rao, G. (2019): "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools," *American Economic Review*, 109, 774–809.

Rehavi, M. M. and S. B. Starr (2014): "Racial Disparity in Federal Criminal Sentences," *Journal of Political Economy*, 122, 1320–1354.

Rose, E. K. (2022): "A Constructivist Perspective on Empirical Discrimination Research," *Working Paper*.

Ross, L., D. Greene, and P. House (1977): "The "False Consensus Effect": An Egocentric Bias in Social Perception and Attribution Processes," *Journal of Experimental Social Psychology*, 13, 279–301.

Rothstein, R. (2017): *The Color of Law: A Forgotten History of How Our Government Segregated America*, Liveright Publishing.

Rugh, J. S. and D. S. Massey (2010): "Racial Segregation and the American Foreclosure Crisis," *American Sociological Review*, 75, 629–651.

Sarsons, H. (2019): "Interpreting Signals in the Labor Market: Evidence from Medical Referrals," *Working Paper*.

Sen, M. and O. Wasow (2016): "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics," *Annual Review of Political Science*, 19, 499–522.

Stoll, M. A. (2009): "Ex-offenders, criminal background checks, and racial consequences in the labor market," *U. Chi. Legal F.*, 381.

Welch, F. (1973): "Black-White Differences in Returns to Schooling," *American Economic Review*, 63, 893–907.

Yinger, J. (1995): *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*, Russell Sage Foundation.

Zafar, M. B., I. Valera, M. Gomez Rodriguez, and K. Gummadi (2017): "Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment," *Proceedings of the 26th International Conference on World Wide Web*.

A Empirical Decompositions

This appendix details our experimental decompositions of total discrimination in Hiring Manager actions into direct and systemic components, following Equations (4) to (6). For $y^0 \in \{2, 3, 4, 5, 6\}$, total discrimination is given by

$$\Delta(y^0) = E[A_i^H \mid G_i = m, Y_i^0 = y] - E[A_i^H \mid G_i = f, Y_i^0 = y]$$

where A_i^H is the Hiring Manager action for Worker i, G_i is Worker i's group (either male m or female f), and Y_i^0 is Worker i's qualification (task-A performance). We estimate total discrimination by the corresponding sample average differences, $\widehat{\Delta}(y)$.

Average direct discrimination at Hiring Manager signal realization $s^H \in \mathcal{S}^H$, where $\mathcal{S}^H = \{1, \dots, 10\}$ in the first experiment and either $\mathcal{S}^H = \{2, 3, 4, 5\}$ or $\mathcal{S}^H = \emptyset$ in the second experiment, is given by

$$\tau(s^H) = E[A_i^H \mid G_i = m, S_i^H = s^H] - E[A_i^H \mid G_i = f, S_i^H = s^H],$$

We again estimate this by the corresponding sample average differences $\hat{\tau}(s^H)$. For the first term of Eq. (4) we then compute average direct discrimination as

$$\hat{E}[\tau(S_i^H) \mid G_i = m, Y_i^0 = y^0] = \frac{1}{N_{m,y^0}} \sum_{i:G_i=m, Y_i^0=y^0} \hat{\tau}(S_i^H),$$

for each y^0, where N_{g,y^0} gives the number of Workers with $G_i = g$ and $Y_i^0 = y^0$. This gives our estimates of average direct discrimination for Equation (4) in Table 2. Estimates of systemic discrimination are then given by

$$\hat{\delta}(f, y^0) = \widehat{\Delta}(y^0) - \hat{E}[\tau(S_i^H) \mid G_i = m, Y_i^0 = y^0]$$

Similar computations yield the estimates of average direct and systemic discrimination in Equations (5) and (6). For the former, average direct discrimination is estimated as

$$\hat{E}[\tau(S_i^H) \mid G_i = f, Y_i^0 = y^0] = \frac{1}{N_{f,y^0}} \sum_{i:G_i=f, Y_i^0=y^0} \hat{\tau}(S_i^H),$$

with systemic discrimination estimated as $\widehat{\delta}(m, y^0) = \hat{\Delta}(y^0) - \hat{E}[\tau(S_i^H) \mid G_i = f, Y_i^0 = y^0]$. For Equation (6) we take an unweighted average of the average direct and systemic discrimination estimates in Equations (4) and (5) to estimate $\bar{\tau}(y^0)$ and $\bar{\delta}(y^0)$, respectively. We average the y^0-specific decompositions by the equal-weighted average of sample group-m and group-f distributions of Y_i^0.

B Additional Examples

B.1 Accurate Statistical Discrimination with Social Information

In this example, we show how accurate statistical discrimination in an initial decision leads to persistent systemic discrimination. This systemic discrimination stems from inflationary signals, which arise endogenously from the social learning and persist in all subsequent decisions. In contrast, if the signaling technology were exogenous, such systemic discrimination would not arise.

Suppose a Worker's productivity Y_i^* is distributed normally with a group-specific mean and common variance: $Y_i \mid \{G_i = g\} \sim N(\mu_g, 1)$ for $\mu_w > \mu_b$. A sequence of evaluators $t = 1, 2, ...$ predict each Worker's productivity with a forecast $A_{it} \in \mathbb{R}$. Before reporting her forecast, evaluator t observes the history of past forecasts $H_{it} = \{A_{i1}, ..., A_{i,t-1}\}$, with $H_{i1} = \emptyset$, and a new signal $\tilde{S}_{it} = Y_i^* + \varepsilon_{it}$, where $\varepsilon_{it} \mid H_{it}, G_i \sim N(0, 1)$. All evaluators have correct knowledge of the distribution of productivity and the signal-generating process.[50] They use Bayes' rule to form a forecast from $S_{it} = \{H_{it}, \tilde{S}_{it}\}$. The researcher selects qualification $Y_i^0 = Y_i^*$ to measure discrimination among equally-productive Workers.

The first evaluator's forecast exhibits direct discrimination, due to accurate statistical discrimination. Namely, she observes a signal of $\tilde{S}_{i1}$ and reports a forecast of $A_1(g, S_{i1}) = (\mu_g + \tilde{S}_{i1})/2$ for a Worker of gender g. Thus there is direct discrimination of $(\mu_w - \mu_b)/2 > 0$. There is no systemic discrimination, because conditional on productivity the signal process is the same for group-w and group-b Workers. Therefore, total discrimination is equal to direct discrimination for the first forecast.

In all subsequent forecasts, however, there is no direct discrimination. The second evaluator reports a forecast of $A_2(g, S_{i2}) = (2A_{i1} + \tilde{S}_{i2})/3$ for a Worker of gender g. Therefore, Workers with the same forecast history and current signal receive the same forecast, regardless of their group. The same is true in subsequent periods: $A_t(g, S_{it}) = (tA_{i,t-1} + \tilde{S}_{it})/(t+1)$ for $t > 1$, which does not depend on g. Intuitively, the Worker's history is a sufficient statistic for her productivity (more formally, the group mean difference in productivity), such that, conditional on the history, there is no information gained from G_i after the initial forecast.

Nevertheless, there is systemic (and therefore, total) discrimination in all forecasts after the first. In the second period, $E[A_2(g, S_{i2}) \mid G_i = w, Y_i^*] = (\mu_w + 2Y_i^*)/3 > (\mu_b + 2Y_i^*)/3 = E[A_2(g, S_{i2}) \mid G_i = b, Y_i^*]$, so systemic discrimination is given by $(\mu_w - \mu_b)/3 > 0$. Similarly, systemic discrimination persists in subsequent periods: in period t, $E[A_t(g, S_{it})|G_i = w, Y_i^*] = (\mu_w + tY^*)/(t+1) > (\mu_b + tY^*)/(t+1) = E[A_t(g, S_{it})|G_i = b, Y_i^*]$, yielding systemic discrimination of $(\mu_w - \mu_b)/(t+1)$.[51] Intuitively, the initial accurate statistical discrimination

[50] Correct knowledge simplifies exposition but is immaterial for this example; all that matters is all evaluators have the same beliefs and this is common knowledge.

[51] In this simple example, systemic discrimination decays to zero as $t \to \infty$, since the forecasts converge

from the first round persists in the signal history, even though there is no new differential updating by group. Given that there is no direct discrimination after the first round, total discrimination is equal to systemic discrimination for the second and subsequent forecasts.

Social learning is a crucial driver of systemic discrimination in this example. If, instead, exogenous signals were directly observable by evaluators (i.e., $H_{it} = \{\tilde{S}_{i1}, ..., \tilde{S}_{i,t-1}\}$), then there would continue to be direct discrimination in each round but there would be no scope for systemic discrimination.

C Institutional Discrimination

C.1 Setup

Consider a set of managers $\mathcal{J}$ at a firm, where each manager $j \in \mathcal{J}$ evaluates a set of candidate workers for a particular task. Each worker i has an observable group identity $G_i \in \{b, w\}$, an *ex ante* unobservable productivity $Y_i^* \in \mathcal{Y}^*$, and a vector of attributes $S_i \in \mathcal{S}$, which is observed by the manager. We write S_i without a j subscript, but in principle signals could be manager-specific. Formally, S_i may contain elements that are observed by some managers and not others. After observing G_i and S_i, manager j takes a scalar action $A_{ij} \in \mathcal{A}$. We abstract from complementarities across workers and other realistic features of labor markets for simplicity; our analysis considers G_i, Y_i^*, S_i, A_{ij}, and Y_i^0 as *iid* random variables with some joint distribution. As in Section 3.1, we take a reduced-form approach to defining the managers action rule: function $A_j(g, s)$ determines manager j's optimal action for a worker with group identity g and realized signal s, such that $A_{ij} = A_j(G_i, S_i)$. We refer to managers with different $A_j(g, s)$ as being of different "types." Different manager types may stem from different combinations of preferences and beliefs, as discussed in Section 4.

To distinguish between individual (manager) behavior and aggregate (institutional) behavior, we consider a firm consisting of a set of managers of potentially different types. For simplicity, we assume each manager in the firm faces the same population of potential workers for the same task (i.e., the same distribution of (G_i, Y_i^*, S_i)) with the same measure of productivity Y_i^*. We define the action rule of the firm $\alpha(g, s)$ as the average rule of its managers: $\alpha(g, s) \equiv \sum_{j \in \mathcal{J}} \pi_j A_j(g, s)$, where π_j denotes the share of workers evaluated by manager j. This allows us to formalize a notion of institutional discrimination as distinct from individual discrimination, along with additional sources of such discrimination.

Definition 4 (Institutional Direct Discrimination)**.** *The firm's actions exhibit institutional direct discrimination at $s \in \mathcal{S}$ if $\alpha(w, s) \neq \alpha(b, s)$.*

to true productivity. But this need not be the case if, for example, signal precision worsens as $t \to \infty$ (e.g., if information acquisition is costly and managers acquire less information as beliefs become more precise). Systemic discrimination may also persist when the initial accurate statistical discrimination is due to differences in signal precision across group, instead of differences in average productivity (see Section 4.4.1).

Let $\mu_j^g(\cdot\ ; y^0)$ denote the distribution of manager j's actions among workers of group g with qualification y^0, and let $M^g(\cdot\ ; y^0) \equiv \sum_{j\in\mathcal{J}} \pi_j \mu_j^g(\cdot\ ; y^0)$ denote the analogous firm distribution.

Definition 5 (Institutional Total Discrimination)**.** *The firm exhibits total discrimination at qualification y^0 if $M^b(a; y^0) \neq M^w(a; y^0)$ for some measurable set of actions $a \subset \mathcal{A}$, and the firm exhibits total discrimination if there exists a y^0 where this holds.*

Define a counterfactual firm action distribution for group g as $\tilde{M}^g(\cdot\ ; y^0) \equiv \sum_{j\in\mathcal{J}} \pi_j \tilde{\mu}_j^g(\cdot\ ; y^0)$.

Definition 6 (Institutional Systemic Discrimination)**.** *The firm exhibits systemic discrimination at qualification y^0 if $M^w(a; y^0) \neq \tilde{M}^b(a; y^0)$ or $M^b(a; y^0) \neq \tilde{M}^w(a; y^0)$ for some measurable set of actions $a \subset \mathcal{A}$, and the manager exhibits systemic discrimination if there exists a y^0 where this holds.*

C.2 Sources of Institutional Discrimination

The composition of managers within the firm—specifically, the distribution of managers' preferences, beliefs, and signaling technologies—play a key role in determining whether discrimination at the individual level translates into institutional discrimination. Formally, the payoff function $u_j(a, y, g)$, the subjective belief about productivity $\hat{F}_{y,j}^g(y)$ and the signaling technology $\sigma_j^g(\cdot\ ; y, y^0)$ can all vary by manager. The first two components determine how individual action rules aggregate to a firm-level action rule, and the final component determines the firm-level signaling technology $\sigma^g(\cdot\ ; y, y^0) \equiv \sum_{j\in\mathcal{J}} \pi_j \sigma_j^g(\cdot\ ; y, y^0)$.

In the case of direct discrimination, different preferences and beliefs lead to different levels of individual direct discrimination. Therefore, managerial composition impacts how individual direct discrimination aggregates to institutional direct discrimination. For example, if managers are divided by the same group identity as workers and favor workers from their own group (i.e., taste-based discrimination stemming from in-group bias), then whether or not institutional direct discrimination arises from individual direct discrimination will crucially depend on which group is dominant at the managerial level.[52] If group-w managers are over-represented relative to group-w workers, then the firm will tend to exhibit institutional direct discrimination against group-b workers. In contrast, with proportional representation and evaluation, such institutional direct discrimination will not arise even if direct discrimination occurs at the individual level.

Institutional systemic discrimination arises from the same two forces as individual systemic discrimination: namely, the functional dependence of the firm's action rule $\alpha(g, s)$ on s and how the signal depends on group identity conditional on qualification. The composition

[52] For example, Antonovics and Knight (2009) show that police officers are more likely to conduct a search if their race differs from that of the driver. Fisman, Paravisini, and Vig (2017) demonstrate that cultural proximity between a loan officer and applicant increases favorable treatment.

of managers determines how the firm-level action rule depends on s. For example, if some managers place weight on an uninformative signal correlated with group membership and others do not, then managerial composition will determine the extent to which the firm's action rule depends on the signal. When the signaling technology differs by manager, the composition of managers also determines the firm-level signaling technology. For example, in Benson, Board, and Meyer-ter Vehn (2019), managers more accurately screen workers with whom they share the same race/ethnicity. Therefore, whether or not institutional systemic discrimination arises from individual systemic discrimination again depends on whether one group is dominant at the managerial level.

Institutional total discrimination also depends on manager composition. For example, in Section 2, suppose that the platform consists of a panel of managers who decide how to reward workers. Aware managers select an action rule that (through its dependence on g) reverses the bias arising from the direct discrimination by consumers while unaware managers select a race-blind action rule. Therefore, whether the actions of a platform composed of such managers exhibits total discrimination will depend on the share of managers that are aware versus unaware of the consumers' bias.

When the signaling technology or productivity distribution is linked to decisions in other markets, then the composition of managers in other markets is also relevant for both individual and institutional systemic discrimination in the current task through its impact on these distributions. For example, in the modified setup of Section 2 where the platform consists of a panel of managers, heterogeneity with respect to the extent of the consumers' bias will impact the managers' signaling technology, and hence the extent of systemic discrimination. Systemic discrimination can arise from individual direct discrimination in other markets even when this individual discrimination does not aggregate to institutional direct discrimination in the other market.