Systemic Discrimination: Theory and Measurement*

J. Aislinn Bohren† Peter Hull‡ Alex Imas§

November 16, 2022

Economics tends to define and measure discrimination as disparities stemming from the direct effects of protected group membership. But work in other fields notes that such measures are incomplete, as they can miss important systemic (indirect) channels. For example, racial disparities in criminal records due to discrimination in policing can lead to disparate outcomes for equally-qualified job applicants despite a race-neutral hiring rule. We develop new tools for modeling and measuring such systemic forms of discrimination. We formalize systemic discrimination as disparities arising from differences in *non-group* characteristics, such as criminal records, among equally-qualified individuals. Systemic disparities can arise both from differences in signaling technologies and differences in opportunities for skill development. Standard tools for measuring direct discrimination, such as audit or correspondence studies, cannot detect systemic discrimination. Instead, we propose a measure based on a decomposition of total discrimination—disparities among equally-qualified individuals—into direct and systemic components. To bring this decomposition to data, we develop a novel *Iterated Audit* experimental paradigm and apply it in a series of hiring experiments and a lab-in-the-field study using real hiring managers. Our findings highlight how discrimination in one domain, due to either accurate beliefs or bias, can drive persistent disparities through systemic channels—even when direct discrimination is eliminated.

*We thank Jason Abaluck, David Arnold, Marianne Bertrand, Leonardo Bursztyn, Damon Jones, Diag Davenport, Will Dobbie, Sam Hirshman, Larry Katz, Matthew Knepper, Matthew Notowidigdo, Matt Lowe, Anna Gifty Opoku-Agyeman, Jane Risen, Heather Sarsons, Fabio Tufano, Damián Vergara, and seminar participants and conference participants at various institutions for helpful comments. Christy Kang and Cuimin Ba provided expert research assistance.

†Department of Economics, University of Pennsylvania: abohren@sas.upenn.edu

‡Department of Economics, Brown University: peter_hull@brown.edu

§Booth School of Business, University of Chicago: 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 group-based disparities tend to focus on the possibility of *direct discrimination*: differential treatment on the basis of the protected characteristic itself, holding other characteristics fixed. 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 2020)—have been the primary theoretical tools for studying the drivers of discrimination. 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, holding other observable characteristics fixed.[2]

A large body of work across many fields, however, takes a broader view of discrimination. Scholars of sociology and the law have long examined disparities through a systems-based approach, in which group-based treatment is seen as a cumulative outcome of both direct and indirect interactions between outcomes and evaluations 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). Computer scientists have more recently shown 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 may fail to capture the full scope of inequity: non-group characteristics may themselves embed 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. Unless the manager considers and adjusts for the recruiter's bias, seemingly non-discriminatory (even

[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).

[2]This includes both experimental methods such as audit and correspondence studies (for review, see Bertrand and Duflo (2016)) and non-experimental methods such as certain outcome-based tests (Knowles, Persico, and Todd 2001; Canay, Mogstad, and Mountjoy 2020).

gender-neutral) promotion rules will tend to lead to worse outcomes for female workers. That is, even if the manager does not *directly* discriminate against female workers conditional on their work histories, female workers will be disadvantaged because they have systematically lower salaries. Such *systemic* discrimination is due to gender-based differences in the non-gender salary characteristic, conditional on the workers' initial qualifications.

A real-world example 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 any within-company transfer was discriminatory because it disadvantaged Black employees who were otherwise qualified but lacked a degree, in part due to existing discriminatory policies 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 at Duke Power. Standard economic measures that condition on observables like educational background would therefore 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). Standard economic 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 systemic discrimination. We first develop a simple theoretical framework to distinguish *direct* discrimination—differential treatment on the basis of group identity itself—and *systemic* discrimination: unwarranted treatment disparities arising indirectly through non-group characteristics, 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 even a constant), thereby accounting for any systemic factors affecting the worker's current productivity itself.

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

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

biases in preferences and beliefs. Institutional direct discrimination is generated through the aggregation of individual direct discrimination. Systemic discrimination can arise from disparities in the interactions of individuals or institutions over time, or across different domains within the same time period.

Formally, we 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 (of, e.g., 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 be driven by other properties of the signal generating process such as group-based disparities in informativeness due to screening actions. *Technological* systemic discrimination arises from differences in the relevant productivity measure itself, for example because of differences in opportunities for human capital development. We illustrate these drivers in a series of theoretical applications, showing 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.[4] Direct discrimination is identified by a conventional audit or correspondence study which measures the causal effect of perceived group membership on a decision, 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 propose a general experimental approach, termed an *Iterated Audit* (IA), which can be used to identify systemic discrimination from our decomposition of total discrimination 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 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

[4] 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 an audit or correspondence study measure of 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. Here, Recruiters made wage offers which were then passed along as a (potentially biased) signal to Managers. The second study explored systemic discrimination due to differences in signal informativeness. Here, Recruiters made binary hiring decisions, and Managers only observed objective signals of productivity if the worker was hired by a Recruiter in the past.

Both 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 also 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 only slightly 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 vast 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 were less likely to observe the performance signals of a female Worker than an equally qualified male, leading to subsequent disparities. Moreover, as predicted by our theory, this systemic discrimination was concentrated among high-performing women who would benefit more from informative signals. Together, these findings show how standard measures of discrimination that *condition* on non-group characteristics would miss large and heterogeneous sources of inequity.

Our third lab-in-the-field experiment used the general 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 for an entry-level job. 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, evaluators were informed of the Pager (2003) study and its connection to the task at hand prior to making their decisions.

Comparing evaluations of the two endogenous sets of resumes identifies total discrimination, while comparing white-endogenous and Black-exogenous resumes—the standard comparison in correspondence studies—gives 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 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. 2020), 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 review related literatures on systemic and direct discrimination. In Section 2 we present a simple motivating example with both forms of discrimination. 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 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.[5] 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).[6]

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)*.[7] 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 polices or practices motivated by a discriminatory purpose. Typically, proof of discriminatory intent is required for the finding of disparate treatment.[8]

A systemic perspective is also often found in the recent literature on algorithmic unfairness (e.g. Angwin et al. 2016; Hardt, Price, and Srebro 2016; Zafar, Valera, Gomez Ro-

[5]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.

[6]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.

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

[8]Landmark cases here include *Washington v. Davis (1976)* and *McClesky v. Kemp (1987)*.

driguez, and Gummadi 2017; Berk, Heidari, Jabbari, Kearns, and Roth 2018). As noted above, 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.[9] 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. 2020; 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 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.

[9] Notable exceptions to the typical focus on direct discrimination in economics include Neal and Johnson (1996), Coate and Loury (1993), Glover, Pallais, and Pariente (2017), Cook (2014), and Sarsons (2019).

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).[10] A parallel empirical literature has developed and applied tools for distinguishing different economic theories of discrimination. Recent advances involve outcome-based tests of racial bias, in both observational (Knowles et al. 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 indirect (systemic) forms of discrimination. Additionally, our framework has implications for the interpretation of group-based disparities that have been documented in the economics literature. 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 systemic discrimination when conditioning on underlying qualifications.[11]

2 A Motivating Example: Signal Inflation

We begin our analysis with a simple example, which illustrates how systemic discrimination can emerge in a two-stage employment decision. Suppose a firm is deciding on a wage offer for a worker with observable group $G \in \{m, f\}$. The worker's unobserved productivity $Y^* \in \mathbb{R}$ is first predicted by a recruiter at the firm. A hiring manager observes this prediction and offers the worker a wage. Formally, the recruiter observes a signal $S^R \in \mathbb{R}$ which is normalized to capture the worker's expected productivity: $E[Y^*|S^R = s^R] = s^R$. For simplicity here we assume that (Y^*, S^R) is independent of G. The recruiter submits a productivity forecast $A^R \in \mathbb{R}$ to the hiring manager after observing G and S^R. The hiring manager observes this forecast as her signal, $S^H = A^R$, and offers the worker a wage $A^H \in \mathbb{R}$.

Suppose the recruiter exhibits *direct discrimination* against group-f workers: for any given signal realization s^R, he reports a higher forecast when $G = m$ than when $G = f$.

[10] 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.

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

Specifically, suppose he reports an accurate forecast of $A^R(f, s^R) = s^R$ for a group-f worker with signal s^R and an inflated forecast of $A^R(m, s^R) = s^R + 1$ for a group-m worker with the same signal. This is the definition of discrimination most often used in economics.[12]

The hiring manager does not have any inherent bias against group-f workers: she seeks to offer a wage equal to expected productivity. If she observed S^R herself, she would offer the worker a wage equal to this accurate productivity signal. However, since she instead relies on the recruiter's forecast, her prediction of the worker's productivity (and thus the wage A^H) depends both on S^H and on her belief about how the recruiter forms it.

First suppose the hiring manager fails to account for the bias of the recruiter: she takes his forecast at face value and offers a wage of $A^H(f, s^H) = A^H(m, s^H) = s^H$ after observing forecast $S^H = s^H$. This decision rule is "neutral," in that it is the same for group-m and group-f workers. Therefore, the hiring manager's actions do not exhibit direct discrimination: a group-m worker and group-f worker with the same signal s^H are given the same wage. However, conditional on the *recruiter* signal s^R, and therefore expected productivity, a group-m worker receives a one unit higher expected wage than a group-f worker: $E[A^H(G, S^H) \mid G = m, S^R = s^R] = s^R + 1$ versus $E[A^H(G, S^H) \mid G = f, S^R = s^R] = s^R$. This is because, conditional on the same signal s^R, the observed forecast S^H (and thus the offered wage A^H), depends on the worker's group.[13] Therefore, although the hiring manager treats all workers with the same forecast (signal) equally, she treats workers with the same expected productivity differently.

This example motivates a broader notion of discrimination, which captures systemic disparities in actions A^H that stem indirectly from the dependence of non-group signal S^H on group identity G: i.e., variables that end up being correlated with group identity through individuals' interactions across multiple markets and domains. We refer to this indirect channel as *systemic discrimination*. Systemic discrimination contrasts with direct discrimination in the action rule which conditions on S^H, i.e. the difference between $A^H(m, s^H)$ and $A^H(f, s^H)$. As we formalize below, it instead corresponds to the difference between $E[A^H(g, S^H) \mid G = m, S^R = s^R]$ and $E[A^H(g, S^H) \mid G = f, S^R = s^R]$, where g is fixed in the action rule to net out direct effects. Here systemic discrimination by the hiring manager arises from the direct discrimination by the recruiter, which results in the hiring manager observing a systematically higher forecast for a group-m worker relative to a group-f worker with the same signal. Note that the extent of systemic discrimination depends on the failure of the hiring manager to account for this direct discrimination when interpreting the forecast.

[12] Bias, either in the form of taste-based discrimination or inaccurate statistical discrimination due to stereotypes, e.g. Bordalo et al. (2019), can generate such a decision rule. Here it cannot be generated by accurate statistical discrimination, since we assume the recruiter's signal and worker productivity are jointly independent of worker group.

[13] Specifically, fixing $S^R = s^R$, $S^H = s^R + 1$ when $G = m$ and $S^H = s^R$ when $G = f$.

This simple model highlights a potential channel for discrimination within our broader definition: when a signal is endogenously generated, in that it depends on the preferences and beliefs of other evaluators (e.g. a recommendation letter or rating), then a manager can still exhibit systemic discrimination *even if* her own beliefs or preferences do not directly favor one group of workers. Given the rich psychology and economics literatures demonstrating 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 (Andre 2022; Brownback and Kuhn 2019), it is plausible that initial biases or stereotypes will lead to persistent disparities even when subsequent evaluations are facially neutral. 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 prior evaluations in a way that obscures its persistent impact.

Our notion of *total discrimination* combines the direct and systemic channels. Formally, it corresponds to the difference between $E[A^H(G, S^H) \mid G = m, S^R = s^R]$ and $E[A^H(G, S^H) \mid G = f, S^R = s^R]$, where the first argument of the manager's decision rule is no longer fixed at g. Here total and systemic discrimination coincide, since the hiring manager does not exhibit direct discrimination. But this is not always the case, as we next illustrate.

Suppose now that the hiring manager is aware of the recruiter's bias and accounts for it when interpreting forecasts: she offers wages $A^H(f, s^H) = s^H$ and $A^H(m, s^H) = s^H - 1$ to undo the inflation in group-m forecasts. In this case, the hiring manager exhibits direct discrimination against group-m workers: conditional on the same forecast, she offers a one unit higher wage to a group-f worker relative to a group-m worker. As in the previous case, the recruiter's direct discrimination translates into systemic discrimination in manager actions: $E[A^H(m, S_i^H) \mid G_i = m, S_i^R = s^R] = s^R > s^R - 1 = E[A^H(m, S_i^H) \mid G_i = f, S_i^R = s^R]$ and $E[A^H(f, S_i^H) \mid G_i = m, S_i^R = s^R] = s^R + 1 > s^R = E[A^H(f, S_i^H) \mid G_i = f, S_i^R = s^R]$. But now, since the hiring manager's direct discrimination in favor of group-f workers exactly offsets the systemic discrimination against group-f workers, she exhibits no total discrimination. That is, conditional on expected productivity $S^R = s^R$, group-m and group-f workers receive the same wage offer: $E[A^H(G, S^H) \mid G = g, S^R = s^R] = s^R$ for $g \in \{m, f\}$.

From these two cases, we see that whether systemic discrimination translates into total discrimination depends crucially on whether the hiring manager is aware of the recruiter's bias: if the hiring manager is unaware, and takes the forecast at face value, then her wage offers also exhibit total discrimination. In contrast, if she is aware of the bias, then she can engage in direct discrimination in the opposite direction to offset the systemic discrimination, resulting in no total discrimination.

This example provides context through which to interpret 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 group-f workers. For example, in the setting outlined above, if some hiring managers are aware of the recruiter's bias and others are not then on average recruiters directly discriminate against group-f workers while hiring managers reverse and directly discriminate against group-m workers. However, group-f workers face systemic and total discrimination across both time periods.[14]

We note that 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. Differences in the subsequent signaling technology that arise from the social learning are a key driver of 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 extending the previous definitions of 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. We also discuss its potential application to other settings.

3.1 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 $G_i \in \{m, f\}$ and an *ex ante* unobservable productivity $Y_i^* \in \mathcal{Y}^*$. For concreteness G_i can be interpreted as any protected characteristic such as individual i's gender, race, age, or ethnicity. Worker i is also characterized by a vector of attributes $S_i \in \mathcal{S}$ (e.g. educational background, prior evaluations, etc.), which is observed by the manager. This vector plays an informational role in the hiring task: it can be interpreted as a signal of productivity Y_i^*, potentially along with G_i.[15] After observing G_i and S_i, manager j takes a scalar action $A_{ij} \in \mathcal{A}$. 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 workers and other realistic features of labor markets for simplicity; our analysis considers G_i, Y_i^*, S_i, A_{ij}, and

[14] 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.

[15] 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.

Y_i^0 (discussed below) as *iid* random variables with some joint distribution.

Rather than explicitly modeling the manager's decision problem here, we take a reduced-form approach: managers follow some systematic decision rule to determine their action choices from their information set. Formally we assume the existence of a function $A_j(g, s)$ that determines manager j's optimal action given a worker's group identity g and the signal s, such that $A_{ij} = A_j(G_i, S_i)$. Absent restrictions on S_i, the existence of such rules is without conceptual loss. We refer to managers with different $A_j(g, s)$ as being of different "types." 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). This model shows how different manager types may stem from different combinations of preferences and beliefs.

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.

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 $Y_i^0 \in \mathcal{Y}^0$, which captures some reference level of productivity. The payoff-relevant productivity in the hiring task could be the same as this measure of qualification, $Y_i^* = Y^0$, or Y_i^* could arise endogenously from Y_i^0 and the actions of other managers and firms.

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.

The following four non-employment contexts illustrate the generality of this setup:

Lending. Loan officers (managers) at a bank (firm) decide whether to lend to borrowers (workers). Borrowers differ in their ability to pay back the loan Y_i^* if it is originated (A_{ij}). 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. Loan officers observe borrowers' credit scores and income (S_i), which provide information about Y_i^*.

Education. Admissions officers (managers) at a school (firm) decide whether to admit students (workers). Students differ in their academic performance Y_i^* if admitted (A_{ij}). 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. Admissions officers observe test scores and recommendation letters (S_i) which predict Y_i^*.

Healthcare. Doctors (managers) at a hospital (firm) decide whether to test patients (workers) for a treatable disease. Patients differ in the disease outcome Y_i^* that is realized if they are not tested (A_{ij}). Doctors observe blood pressure (S_i), which is informative about Y_i^*. 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. Judges (managers) in a district (firm) decide whether to release defendants (workers) before trial. Defendants differ in their potential for pretrial misconduct Y_i^* that is realized if they are released under some conditions (A_{ij}). 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. Judges observe defendants' prior criminal record (S_i), which provides information about Y_i^*.

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 types of discrimination in the manager's action with respect to worker group G_i: *direct* and *systemic*. Direct discrimination arises causally from the worker's group identity itself, because of manager preferences or beliefs. Systemic discrimination arises from group-based differences in non-group characteristics S_i, which lead to different actions as a function of group identity in the absence of direct (i.e. causal) effects of G_i. Such group-based differences in S_i may stem from direct discrimination in other periods or markets. *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*.

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

Definition 1 (Direct Discrimination). *Manager j's actions exhibit individual direct discrimination if $A_j(m,s) \neq A_j(f,s)$ for some $s \in \mathcal{S}$. The firm's actions exhibit institutional direct discrimination if $\alpha(m,s) \neq \alpha(f,s)$ for some $s \in \mathcal{S}$.*

Because of the conditioning on all relevant non-group characteristics S_i, direct discrimination is a causal concept: it follows from the structure of the action rules $A_j(g,s)$ and $\alpha(g,s)$, in particular their functional dependence on worker group membership g. 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 by managers—what we term individual direct discrimination—arising from causal effects of group membership on the manager's preferences or beliefs about productivity. We discuss these canonical sources of direct discrimination in Section 4. In Section 5, we discuss how direct discrimination can be measured by audit or correspondence studies, which measure the causal effect of G_i by randomizing over or conditioning on the non-group characteristics S_i.[16]

Our definition of systemic discrimination departs from these economic models by considering the non-causal dependence between manager or firm actions and worker group, conditional on worker qualification:

Definition 2 (Systemic Discrimination). *Manager j's actions exhibit individual systemic discrimination if $A_j(g,S_i)$ is not independent of G_i conditional on Y_i^0 for some $g \in \{m,f\}$. The firm's actions exhibit institutional systemic discrimination if $\alpha(g,S_i)$ is not independent of G_i conditional on Y_i^0 for some $g \in \{m,f\}$.*

Because this definition fixes worker group membership g in the action rules, systemic discrimination is unaffected by any direct effect of group identity on manager or firm actions. Instead, it arises from the statistical relationship between non-group characteristics S_i and group identity G_i in the population of workers. We condition this relationship on Y_i^0, such that systemic discrimination only arises among equally "qualified" workers with different non-group characteristics. For example, a word-of-mouth recruitment practice that prioritizes workers with a social connection to the firm may be systemically discriminatory when men are more connected than equally qualified women (perhaps because of past direct discrimination in hiring). The practice of "redlining" in mortgage markets is another example: borrowers from majority-white neighborhoods (as recorded in S_i) may be prioritized for a

[16] Here and below we abstract away from several conceptual issues with studies that manipulate signals of protected characteristics, such as worker names, instead of the perceived characteristics 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), and 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 present issues for measurement error and interpretation of disparities as direct discrimination by animus versus statistical discrimination.

loan over borrowers from majority-Black neighborhoods, regardless of borrower's race G_i. If such treatment differences remain conditional on the relevant measure of qualification Y_i^0, then $A_j(g, S_i)$ and $\alpha(g, S_i)$ will be conditionally correlated with G_i.

Definition 2 aligns broadly with literatures considering systemic (or structural) discrimination as a form of inequality operating indirectly through non-group characteristics (e.g. those reviewed in Section 1.1). As this work outlines, such discrimination can emerge when systems (or components of a system) either interact across time (i.e. "past-in-present" discrimination) or interact contemporaneously across different domains (i.e. "side-effect" discrimination).[17] Both forms may emerge even when managers in the current task exhibit no direct discrimination, if they fail to account for discrimination in the past or in other domains.[18] The literature also discusses how such 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 non-group characteristics of this group.[19]

Analogous to the case of direct discrimination, different underlying sources can give rise to systemic discrimination. We define and discuss two key sources—the signaling technology and the productivity distribution conditional on qualification—in Section 4. Group differences in these sources can arise endogenously from direct discrimination in other markets as well as from design choices in the present market.

Total discrimination—the overall dependence between manager or firm actions and worker group, conditional on worker qualification—combines these direct and systemic channels:

Definition 3 (Total Discrimination)**.** *Manager j's actions exhibit individual (total) discrimination if $A_j(G_i, S_i)$ is not independent of G_i conditional on Y_i^0. The firm's actions exhibit institutional (total) discrimination if $\alpha(G_i, S_i)$ is not independent of G_i conditional on Y_i^0.*

Total discrimination can arise from direct (i.e. causal) effects of the group on manager actions or from systemic discrimination through non-group characteristics.

In Section 4.5, we develop several additional examples to illustrate different ways sys-

[17] 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.

[18] 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).

[19] 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 that involving a height-to-weight threshold calibrated using Anglo-Saxon data, which led to the disproportionate rejection of South-East Asian applicants.

temic and total discrimination can arise, while in Section 5 we discuss measurement and identification. We bring these definitions to data in Sections 6 and 7.

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 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 non-group characteristics. 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 treatment disparity by group 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.[20]

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 as in Section 2, setting Y_i^0 to the screening test performance (e.g. S_i^R) measures systemic discrimination in hiring manager actions that stems from direct discrimination by the recruiter. Similarly,

[20] 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.

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 4.5.2 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 or firm'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.[21] Total discrimination with respect to qualification $Y_i^0 = Y_i^*$ encompasses this case, whereas direct discrimination does not.[22] This case aligns total discrimination with the legal notion of disparate impact, as it allows for disparities relevant to "business necessity."[23]

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.

4 Sources of Discrimination

We now explore and contrast potential sources of direct and systemic discrimination, as defined above. To do so we microfound the reduced-form action rule in terms of a manager's preferences and beliefs, and delineate how the relationship between the signal, productivity, and qualification can vary by group. We then discuss sources of individual direct and systemic discrimination, followed by a discussion of sources of institutional discrimination. Finally, we outline several additional theoretical applications to illustrate the different sources.

[21] 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).

[22] Alternatively, certain non-group characteristics may enter the manager or firm'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.

[23] 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_{ij} is a prediction of some latent state Y_i^* or an algorithmic recommendation based on such a prediction (Berk et al. 2018; Arnold, Dobbie, and Hull 2021).

4.1 Setup

We first develop a single manager's decision problem, suppressing the j subscript to ease notation. 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 $g \in \{m, f\}$. 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. 2020).

Specifically, the manager holds subjective belief $\hat{F}_y(y|g)$ about the distribution of productivity for group g, which we refer to as the perceived productivity distribution, and subjective belief $\hat{F}_s(s|y, g)$ about the signal distribution for a worker from group g with productivity y. We refer to subjective beliefs about the signal generating process as the perceived signaling technology. Given these subjective distributions, the manager uses Bayes' rule to form a posterior belief $\hat{F}_y(y|s, g)$ about the worker's productivity after observing signal realization s. She chooses an action to maximize expected utility with respect to this posterior belief:

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

which yields the reduced-form decision rule introduced in Section 3.1.

Only beliefs about the productivity distribution and signaling technology are relevant for the manager's decision—and hence, are the only relevant sources for direct discrimination. In contrast, the *true* productivity distribution and signaling technology are relevant for capturing sources of systemic discrimination. Let $F_y(y|y^0, g)$ denote the conditional productivity distribution for workers with qualification $Y_i^0 = y^0$ and group identity $G_i = g$. Let $F_s(s|y, y^0, g)$ denote the conditional signaling technology for workers with productivity $Y_i^* = y$, qualification y^0 and group identity g. From these distributions, as well as the qualification distribution $F_0(y^0|g)$, we construct the true (unconditional) productivity distribution and signaling technology respectively denoted by $F_y(y|g)$ and $F_s(s|y, g)$. From Bayes' rule, we can analogously derive the posterior belief $F_y(y|s, g)$ about a worker's productivity conditional on observing signal realization s.

4.2 Sources of Direct Discrimination

Individual direct discrimination arises when the manager's action rule depends on group identity. This dependence stems from either the manager's preferences or beliefs. In the case of classic (i.e. accurate) statistical discrimination, the channel is beliefs. The manager has an accurate posterior belief about productivity that takes group membership into account, $\hat{F}_y(y|s, g) = F_y(y|s, g)$. The manager's payoffs do not depend on worker group: $u(a, y, m) = u(a, y, f)$. Generally there is direct discrimination when the posterior distribution depends on

g, either because the productivity distribution $F_y(y|g)$ or the signaling technology $F_s(s|y,g)$ differ by group (Phelps 1972; Arrow 1973; Aigner and Cain 1977).

Individual direct discrimination can also arise with deviations from accurate statistical discrimination, which is typically termed "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. 2020), in which the manager has an incorrect posterior belief about the worker's productivity, $\hat{F}_y(y|s,g) \neq F_y(y|s,g)$, which depends on g. 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).[24] Direct discrimination can also arise when the firm constrains the decisions of its managers through various institutional norms and regulations. For example, a firm may require its managers to base decisions on an algorithmic hiring rule that is discriminatory, or employ discriminatory policies such as race-based quotas.

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. Formally, it arises from the functional dependence of $A(g,s)$ on s and how the distribution $F_s(s|y^0,g)$ depends on g. Since $F_s(s|y^0,g)$ is constructed from the conditional signaling technology $F_s(s|y,y^0,g)$ and the conditional productivity distribution $F_y(y|y^0,g)$—specifically, from integrating the product of the corresponding densities over $y \in \mathcal{Y}^*$—there are two channels that can generate systemic discrimination: an informational channel given by group differences in $F_s(s|y,y^0,g)$, and a technological channel given by group differences in $F_y(y|y^0,g)$. 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 $F_s(s|y,y^0,g)$ 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. 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

[24]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 et al. (2020) and Grau and Vergara (2021) for discussions of this issue in economics.

(Y_i^0) may have credit histories (S_i) that are differentially informative due to discrimination in past borrowing opportunities.

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 and higher signal realizations lead to more favorable actions. For example, in the previous criminal justice example, suppose one group is more likely to have a prior criminal offense than the other and that having a prior criminal offense reduces the probability of being released on bail or being considered for an interview (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 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—combined with inaccurate beliefs about the distribution of evaluator types is a key mechanism behind signal inflation. Returning to the criminal justice example, suppose the bail judge believes that there is no direct discrimination in policing, and therefore, having a prior criminal offense reflects the same underlying criminal activity for both groups. But in reality, there is no underlying group-based difference in criminal activity: the differential likelihood of having a prior criminal offense stems from direct discrimination in policing. This inaccurate belief about policing will then lead to systemic discrimination stemming from signal inflation. This channel is illustrated in the motivating Section 2 example and empirically documented in Section 6.1.

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 therefore leads to a larger reduction in uncertainty over productivity for this group, generally leading to systemic discrimination. Unlike signal inflation, the direction of systemic discrimination from screening tends to vary with 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 is set to be 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. Such a difference in precision might arise, for example, when the signal is a test specifically trained to screen workers of group m and which signals the productivity of group f less reliably.[25] Group f may also have less informative productivity signals because they had less previous

[25] 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.

opportunities to establish a record, as in the credit example discussed above. We illustrate this channel in Section 4.5.1 and we document empirically in Section 6.2.

Technological Systemic Discrimination emerges from group-based differences in productivity Y_i^*, conditional on initial qualification Y_i^0. Formally, it is generated when $F_y(y|y^0, g)$ depends on g. This channel is clearly only present 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 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 one group than another relative to Y_i^0. For example, suppose group m is given more access to training and skill development due to discrimination in prior decisions.[26] Technological systemic discrimination can also arise from other properties of the conditional productivity distribution. 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.[27]

We note that group differences in the distribution of worker qualification cannot lead to systemic discrimination with respect to that qualification, as the definition of systemic discrimination conditions on qualification. This observation highlights how the chosen qualification measure is a 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 $F_s(s|y, g)$ and the unconditional productivity distribution $F_y(y|g)$ contribute to systemic discrimination. At the other extreme, when Y_i^0 is set to Y_i^*, only differences in the unconditional signaling technology $F_s(s|y, g)$ contribute to systemic discrimination: differences in $F_y(y|g)$ play no role. In between these extremes, differences in the conditional signaling technology $F_s(s|y, y^0, g)$ and the conditional productivity distribution $F_y(y|y^0, g)$ can both contribute to systemic discrimination. We also note there is no scope for "inaccurate" systemic discrimination: only true distributions contribute to systemic discrimination.[28]

[26] 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.

[27] Analogous to how direct discrimination can arise from omitted payoff bias (see Footnote 24), 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.

[28] Inaccurate beliefs about $F_s(s|y, y^0, g)$ 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 university administrator may perceive the signaling technology for a particular standardized test to be identical across groups, and therefore, choose to continue using it despite discriminatory signal inflation.

Accurate Statistical versus Systemic Discrimination. It is instructive to highlight differences in the sources of accurate statistical (direct) discrimination and systemic discrimination. While both can arise in the absence of biased preferences or beliefs, they differ in how they are driven by group-based differences in the signaling technology and productivity distribution. For example, when $Y_i^0 = Y_i^*$, differences in the signaling technology $F_s(s|y, g)$ can drive both forms of discrimination, but differences in the productivity distribution $F_y(y|g)$ can only lead to accurate statistical discrimination (i.e. there is no technological systemic discrimination). When $Y_i^0 \neq Y_i^*$, differences in the conditional signaling technology $F_s(s|y, y^0, g)$ can still drive both forms of discrimination. But in this case, differences in the conditional productivity distribution $F_y(y|y^0, g)$ can also drive systemic discrimination.

To emphasize the difference between accurate statistical and systemic discrimination, consider a "group-blind" manager whose payoffs and beliefs are *both* unaffected by G_i. Formally, the manager's payoff satisfies $u(a, y, m) = u(a, y, f)$ and his posterior beliefs $\hat{F}_y(y|s, g) = \hat{F}_y(y|s)$ only depend on the non-group signal S_i for $g \in \{m, f\}$. These beliefs may be accurate on average, in that $\hat{F}_y(y|s) = Pr(G_i = m)F_s(y|s, m) + Pr(G_i = f)F_s(y|s, f)$ is equal to the true productivity distribution conditional on the signal. Importantly, the beliefs do not condition on worker group, as in classic accurate statistical discrimination models. Since payoffs and beliefs are independent of worker group, there is no direct discrimination: conditional on S_i, worker group has no effect on the manager's action rule. Yet there will be systemic discrimination when the signals entering this group-blind action rule have a different distribution among equally-qualified group-m and group-f workers.

4.4 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 (reintroducing the manager subscript j) the payoff function $u_j(a, y, g)$, the subjective beliefs $\hat{F}_{y,j}(y|g)$ and $\hat{F}_{s,j}(s|y, g)$, and the signaling technology $F_{s,j}(s|y, y^0, g)$ 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 $\phi_s(s|y, y^0, g) \equiv \sum_{j \in \mathcal{J}} \pi_j F_{s,j}(s|y, y^0, g)$.

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.[29] If group-m managers are over-represented relative to group-m workers, then the firm will tend to exhibit institutional direct discrimination against group-f workers. In contrast, with proportional representation and evaluation, such institutional direct discrimination will not arise even if direct discrimination occurs at the individual level. We illustrate in such compositional effects in Section 4.5.3.

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, $\phi_s(s|y^0, g) = \int_{\mathcal{Y}^*} \phi_s(s|y, y^0, g) dF_y(y|y^0, g)$. The composition 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 $\phi_s(s|y^0, g)$. 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. We illustrate compositional effects for institutional systemic discrimination in Section 4.5.3.

Institutional total discrimination also depends on manager composition. For example, in Section 2, aware hiring managers select an action rule that (through its dependence on g) reverses the bias arising from the direct discrimination by recruiters while unaware hiring managers select a group-blind action rule. Therefore, whether the actions of a firm composed of such managers exhibits total discrimination will depend on the share of managers that are aware versus unaware of the recruiters' 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 $F_s(s|y, y^0, g)$ and/or $F_y(y|y^0, g)$. For example, in the setup of Section 2, heterogeneity with respect to the extent of the Recruiters' bias will impact the Hiring 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 (see the example in Appendix B.2).

[29] 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.

4.5 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 three additional examples to illustrate other sources of systemic discrimination—including screening and direct discrimination in a concurrent decision in another domain—as well as how individual systemic discrimination can lead to institutional systemic discrimination.

4.5.1 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 a standardized 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 is 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.

This 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 in each group, $Y_i^* \mid G_i \sim N(0,1)$, but the manager's signal $S_i = Y_i^* + \varepsilon_i$ has a group-specific precision: $\varepsilon_i \mid Y_i^*, G_i \sim N(0, 1/\eta_{G_i})$ for $\eta_m > \eta_f > 0$, so group m has a more precise productivity signal. The distribution of S_i conditional on $Y_i^* = y$ and $G_i = g$ is $N(y, 1/\eta_g)$ and the posterior expected productivity conditional on $S_i = s$ and $G_i = g$ is $E[Y_i^* | S_i = s, G_i = g] = s\frac{\eta_g}{1+\eta_g}$.

Suppose the manager hires all workers whose posterior expected productivity is at or above some threshold $t \in \mathbb{R}$: $A(g,s) = \mathbb{1}\left[E[Y_i^*|S_i = s, G_i = g] \geq t\right]$. From $E[Y_i^*|S_i = s, G_i = g] = s\frac{\eta_g}{1+\eta_g}$, the manager thus hires workers of group g with $S_i \geq t\frac{1+\eta_g}{\eta_g}$. Group-$f$ workers face a higher signal threshold, since $\frac{1+\eta_f}{\eta_f} > \frac{1+\eta_m}{\eta_m}$. Therefore, there is direct discrimination against group f, stemming from the higher cutoff arising from their less precise productivity signal. Specifically, workers with $S_i \in (t\frac{1+\eta_m}{\eta_m}, t\frac{1+\eta_f}{\eta_f}]$ are hired when $G_i = m$ but not hired when $G_i = f$, while workers with other signals are either hired or not hired regardless of G_i.

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 $g \in \{m, f\}$, systemic discrimination is captured by

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

where $\Phi(\cdot)$ gives the standard normal distribution.[30] Since $\eta_f \neq \eta_m$, 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-m versus group-f workers.

Systemic discrimination in this screening action is heterogeneous across worker productivity levels. With $\eta_m > \eta_f > 0$, the systemic discrimination hurts group-f workers at high levels of productivity (where $y > t\frac{1+\eta_g}{\eta_g}$) and *favors* group-$f$ 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-f workers more likely to have a signal above the effective threshold by chance, while high-productivity group-f 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 f overall. Here, there are fewer high-productivity group-f workers hurt by the higher signal variance than low-productivity group-f workers helped by it. Conversely, in a "lemon-dropping" market with a majority of workers hired ($t < 0$) the systemic discrimination hurts group-f 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

[30]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 $Pr(S_i \geq t\frac{1+\eta_g}{\eta_g}|Y_i^* = y, G_i = g') = Pr(\eta_{g'}(S_i - y) \geq \eta_{g'}(t\frac{1+\eta_g}{\eta_g} - y)|Y_i^* = y, G_i = g') = 1 - \Phi(\eta_{g'}(t\frac{1+\eta_g}{\eta_g} - y))$.

impacts total discrimination in the setting.

4.5.2 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 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 $E[A_1(g, S_i)|Y_i^0 = y, G_i = m] - E[A_1(g, S_i)|Y_i^0 = y, G_i = f] = 1 - \gamma > 0$.[31] 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.

4.5.3 Managerial Composition and Institutional Discrimination

Overview. Our final example illustrates how manager composition impacts institutional discrimination in a setting where managers are divided into two groups, m and f, and favor workers from their own group. This example shows how the distribution of manager types can play a crucial role in determining whether individual direct or systemic discrimination translates into comparable discrimination at the institutional level. It also highlights the difficulty of overcoming systemic discrimination at the institutional level: given direct discrimination either in the past or in a different decision-relevant domain, systemic discrimination will persist for the decision at hand even if the composition of managers is fully representative. Here total discrimination dissipates if and only if the composition of managers is unbalanced in the *opposite* direction of the managerial group that generated the direct discrimination.

Application. Return to the set-up of Section 2, but now suppose there are two types of recruiters and hiring managers who themselves belong to group m or group f. Each manager type exhibits in-group bias towards workers from the same group. Specifically, recruiters in group m use decision rule $A^{R,m}(f, s) = s$ and $A^{R,m}(m, s) = s + 1$, which inflates the productivity forecasts of workers from group m, while recruiters in group f use decision rule $A^{R,f}(f, s) = s + 1$ and $A^{R,f}(m, s) = s$, inflating the forecast of workers from group f. Hiring managers in each group inflate wages in a similar way: $A^{H,m}(f, s) = s$, $A^{H,m}(m, s) = s + 1$,

[31] 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 $g = f$.

$A^{H,f}(f,s) = s+1$ and $A^{H,f}(m,s) = s$. Suppose share $\pi_m^R, \pi_m^H \in [0,1]$ of recruiters and hiring managers are in group m, respectively. Hiring managers are aware of the in-group bias of recruiters and know the share of group-m recruiters, π_m^R. Each type of manager evaluates an equal share of group-m and f workers.

Recruiters and hiring managers both exhibit direct discrimination against the out-group, as evidenced by $A^{R,g}(m,s) \neq A^{R,g}(f,s)$ and $A^{H,g}(m,s) \neq A^{H,g}(f,s)$ for each manager type $g \in \{m, f\}$. Given firm-level decision rules $\alpha^R(f,s) = \pi_m^R s + (1-\pi_m^R)(s+1) = s+1-\pi_m^R$ and $\alpha^R(m,s) = \pi_m^R(s+1) + (1-\pi_m^R)s = s + \pi_m^R$, recruiters exhibit institutional direct discrimination against group f when group m is dominant ($\pi_m^R > 1/2$), and conversely for institutional direct discrimination against group m. The same holds for managers.

Hiring managers do not exhibit systemic discrimination if and only if the distribution of recruiters is balanced: i.e., $\pi_m^R = 1/2$. Otherwise, both group-m and group-f hiring managers exhibit systemic discrimination against workers with the same group identity as the minority recruiter group, due to inflationary signals. Systemic discrimination against these workers by hiring managers from the same group is exactly offset by these hiring managers' direct discrimination favoring these workers, resulting in no total discrimination by these hiring managers. In contrast, systemic discrimination against these workers by hiring managers from the other group compounds their direct discrimination against these workers, resulting in an even larger measure of total discrimination.

Given the share of group-m recruiters, whether there is total discrimination at the institutional level depends on the share of group-m hiring managers. If, for example, all recruiters are in group m ($\pi_m^R = 1$), then there is no total discrimination if and only if all hiring managers are in group f. Therefore, when there is initial imbalance by group for recruiters, total discrimination persists even when there is balance in the next stage. This is because of the systemic discrimination by hiring managers that stems from the imbalance of recruiters; it takes an imbalance of equal magnitude in the opposite direction to overcome.

5 Measuring Systemic Discrimination

We now develop measures of systemic discrimination which leverage novel Kitagawa-Oaxaca-Blinder 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.[32] Total discrimination by manager j at qualification level $y^0 \in \mathcal{Y}^0$ is given by

$$\Delta_j(y^0) \equiv E[A_j(G_i, S_i) \mid G_i = m, Y_i^0 = y^0] - E[A_j(G_i, S_i) \mid G_i = f, Y_i^0 = y^0], \tag{1}$$

with an analogous definition of $\Delta(y^0)$ with respect to $\alpha(g, s)$ for firm-level (institutional) total discrimination. A finding of $\Delta(y^0) > 0$ would mean, for example, that the firm hires group-m workers with qualification y^0 at a higher rate than equally-qualified group-f workers. Correspondingly, direct discrimination by manager j at signal realization $s \in \mathcal{S}$ is given by

$$\tau_j(s) \equiv A_j(m, s) - A_j(f, s), \tag{2}$$

and analogously $\tau(s) \equiv \alpha(m, s) - \alpha(f, s)$ for firm-level (institutional) direct discrimination. A finding of $\tau(s) > 0$ would mean, for example, that belonging to group m vs. f causes workers with non-group characteristics s to be hired more often at the firm. Finally, systemic discrimination by manager j at qualification level $y^0 \in \mathcal{Y}^0$ is given by

$$\delta_j(g, y^0) \equiv E[A_j(g, S_i) \mid G_i = m, Y_i^0 = y^0] - E[A_j(g, S_i) \mid G_i = f, Y_i^0 = y^0], \tag{3}$$

for $g \in \{m, f\}$, with again an analogous definition of $\delta(g, y^0)$ with respect to $\alpha(g, s)$ for firm-level (institutional) systemic discrimination. A finding of $\delta(g, y) > 0$ would capture, for example, the difference in hiring rates among equally-productive group-m and group-f workers that arises indirectly from the non-group characteristics S_i.

5.1 Decomposing Total Discrimination

Our decomposition of total discrimination into direct and systemic components follows directly from Equations (1) to (3). For each manager j, we have:

$$\overbrace{\Delta_j(y^0)}^{\text{Total discrimination}} = \underbrace{E[\tau_j(S_i) \mid G_i = m, Y_i^0 = y^0]}_{\text{Average direct discrimination}} + \underbrace{\delta_j(f, y^0)}_{\text{Systemic discrimination}} \tag{4}$$

by adding and subtracting $E[A_j(f, S_i) \mid G_i = m, Y_i^0 = y^0]$ to and from the definition of $\Delta_j(y^0)$, and rearranging terms.[33] Equation (4) shows that total discrimination at qualification level y^0 for each manager j 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

[32] 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 instead consider mean disparities in the indicators $\mathbb{1}[A_{ij} \leq a]$ for $a \in \mathcal{A}$.

[33] Specifically, $\Delta_j(y^0) = E[A_j(m, S_i) \mid G_i = m, Y_i^0 = y^0] - E[A_j(f, S_i) \mid G_i = f, Y_i^0 = y^0] - E[A_j(f, S_i) \mid G_i = m, Y_i^0 = y^0] + E[A_j(f, S_i) \mid G_i = m, Y_i^0 = y^0] = E[A_j(m, S_i) - A_j(f, S_i) \mid G_i = m, Y_i^0 = y^0] + (E[A_j(f, S_i) \mid G_i = m, Y_i^0 = y^0] - E[A_j(f, S_i) \mid G_i = f, Y_i^0 = y^0])$. The first expectation equals $E[\tau_j(S_i) \mid G_i = m, Y_i^0 = y^0]$ while the second term in parentheses equals $\delta_j(f, y^0)$.

distribution for workers from group m with qualification level y^0, and systemic discrimination at qualification level y^0 when the manager uses the action rule for group f.

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.[34] 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—as we discuss more below.

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

$$\mathit{\Delta}_j(y^0) = E[\tau_j(S_i) \mid G_i = f, Y_i^0 = y^0] + \delta_j(m, y^0) \tag{5}$$

by adding and subtracting $E[A_j(m, S_i) \mid G_i = f, Y_i^0 = y^0]$ to and from the definition of $\mathit{\Delta}_j(y^0)$ and rearranging terms. Equation (5) decomposes total discrimination into average direct discrimination with respect to the signal distribution for workers from group f and systemic discrimination when the firm uses the action rule for group m, all at qualification level y^0. Averaging Equations (4) and (5) yields a third decomposition:

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

where $\bar{\tau}_j(y^0) \equiv \frac{1}{2}(E[\tau_j(S_i) \mid G_i = m, Y_i^0 = y^0] + E[\tau_j(S_i) \mid G_i = f, Y_i^0 = y^0])$ is an unweighted average of the direct discrimination terms in equations Equations (4) and (5), while $\bar{\delta}_j(y^0) \equiv \frac{1}{2}(\delta_j(m, y^0) + \delta_j(f, y^0))$ is an unweighted average of the systemic discrimination terms.[35]

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 these challenges can be overcome, 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: $\mathit{\Delta}_j(y^0) = E[A_{ij} \mid G_i = m, Y_i^0 = y^0] - E[A_{ij} \mid G_i = f, Y_i^0 = y^0]$ for each

[34] See, e.g., Neumark (1988) for a discussion of this approach.

[35] Each of the three decompositions Eqs. (4) to (6) can be derived at the firm level by averaging across managers j. For example, total institutional discrimination can be decomposed as $\mathit{\Delta}(y^0) = \sum_{j \in \mathcal{J}} \pi_j \mathit{\Delta}_j(y^0) = \bar{\tau}(y^0) + \bar{\delta}(y^0)$ where $\bar{\tau}(y^0) = \sum_{j \in \mathcal{J}} \bar{\tau}_j(y^0)$ and $\bar{\delta}(y^0) = \sum_{j \in \mathcal{J}} \bar{\delta}_j(y^0)$ average measures of direct institutional discrimination and systemic institutional discrimination, respectively.

$y^0 \in \mathcal{Y}^0$. 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 $E[A_{ij} \mid G_i = m] - E[A_{ij} \mid G_i = f]$.

To apply our decomposition to the case where Y_i^0 is known, we propose a simple experimental approach to identify the direct and total discrimination components in equations (4)-(6). The systemic discrimination components are then given by subtracting the direct discrimination components from the identified level of total discrimination. 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 indirect discrimination.

The first IA step randomizes manager perceptions of group membership, as in a conventional 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 $A_{ij} = A_j(\tilde{G}_i, S_i)$ and where $\tilde{G}_i \perp\!\!\!\perp (G_i, S_i, Y_i^*) \mid Y_i^0$ by virtue of the randomization. For example, a researcher may take a set of real male and female resumes and randomize distinctively-gendered names among equally-qualified workers, holding fixed all other information on the resume.[36] The researcher then solicits manager actions A_{ij} in the experimental sample. Comparing the response to group-m workers randomized to $\tilde{G}_i = m$ with the response to group-m workers randomized to $\tilde{G}_i = f$, at qualification level y^0, identifies the direct discrimination component of (4):

$$\begin{aligned}
&E[A_{ij} \mid \tilde{G}_i = m, G_i = m, Y_i^0 = y^0] - E[A_{ij} \mid \tilde{G}_i = f, G_i = m, Y_i^0 = y^0] \\
&= E[\tau_j(S_i) \mid G_i = m, Y_i^0 = y^0],
\end{aligned}$$

Similarly, comparing the response to group-f workers randomized to $\tilde{G}_i = m$ with the response to group-f workers randomized to $\tilde{G}_i = f$, 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 who did not have their perceived group membership affected 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 $\Delta_j(y^0) = E[A_{ij} \mid G_i = m, Y_i^0 = y^0] - E[A_{ij} \mid G_i = f, Y_i^0 = y^0]$ identifies one of the three systemic discrimination components in equations (4)-(6).

Figure 1 illustrates an example iterated audit conducted with male (group m) and female

[36] Again, we abstract away from conceptual and econometric issues with randomizing signals of group membership, such as "distinctively Black names," in analyses of direct racial discrimination; see Footnote 16.

FIGURE 1. Iterated Audit Example

(group f) resumes, where the researcher is interested in studying discrimination conditional on worker education Y_i^0. Resumes A and C represent "endogenous" profiles of male and female 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 female candidate with male characteristics, generated by randomizing a distinctively female photo to the real male resume A (holding all other elements, such as education and work experience, fixed). Hiring disparities arising from such randomization can be attributed to 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 (here, female) due to the non-group characteristics (here, work experience).

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, $\tau_j(s) = E[A_{ij} \mid G_i = m, S_i = s] - E[A_{ij} \mid G_i = f, S_i = s]$, 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 $A_{ij} \in \{0, 1\}$ indicates a hiring decision, observed output $Y_{ij} = A_{ij}Y_i^0$ gives a selective measure of qualification: workers who are hired ($A_{ij} = 1$) reveal their

[37]More generally, when S_i is only partially observed, variants of the frameworks of Altonji et al. (2005) and Oster (2017) 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).

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 j 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 $E[A_{ij} \mid G_i = m] - E[A_{ij} \mid G_i = f]$.[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, with the (quasi-)experimental action variation underlying the Arnold et al. (2022) approach. Specifically, consider a set of group-m 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 Arnold et al. (2022) selection correction technique to adjust the experimental disparities $E[A_{ij} \mid \tilde{G}_i = f, G_i = m] - E[A_{ij} \mid \tilde{G}_i = m, G_i = m]$. This term could then be subtracted from the original Arnold et al. (2022) measure of total discrimination to identify 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 (2017) 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.

[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 (which they call disparate impact) uses the fact that Y_i^0 is binary, they also discuss extensions to multivalued or continuous Y_i^0.

Broadly, then, 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 illustrate how our decomposition can be used to measure systemic discrimination in two stylized lab experiments conducted on the *Prolific.co* platform. The first experiment illustrates systemic discrimination arising from signal inflation, as in the motivating example in Section 2. The second study illustrates how systemic discrimination can be heterogeneous in screening decisions, similar to the example in Section 4.5.1. In these studies, a pool of workers face evaluations from two sets of managers. The first set of managers generate initial evaluations of workers based on their group identity and a productivity signal. The second set of managers evaluates workers based on their group identity and an endogenous productivity signal generated by the first set of managers. Worker qualification is chosen so that there is no systemic discrimination in initial evaluations: total discrimination is equal to direct discrimination in this stage. This direct discrimination can lead to systemic discrimination in the second-stage evaluation, alongside any direct discrimination.[39]

6.1 Signal Inflation

Our first lab experiment shows how systemic discrimination can arise from signal inflation. The design closely mirrors the setup of the theoretical example in Section 2. A pool of workers ($N = 100$) completed a two sets of tasks (A and B) and answered a series of demographic questions, including self-reported gender, which we consider the group G_i.

The workers faced evaluations from two sets of managers. Each manager was given 10 experimental units (10 EU=\$1) as an initial budget. Recruiters ($N = 200$) were shown information about two Workers and reported their highest willingness to pay to hire each. Specifically, each Recruiter was shown a signal S_i^R for each assigned Worker consisting of the number of questions the Worker completed correctly on Task A, 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

[39] Preregistration materials for both studies can be found here https://aspredicted.org/TK7_R4J and https://aspredicted.org/K3Q_RPK.

to hire Worker i, between 0 and 10 EUs.[40] Recruiter wage offers were 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.[41] If the worker was not hired, the Recruiter did not pay anything and kept their endowment.[42] Task B performance is thus the relevant measure of Worker productivity in this first stage of the experiment, which we denote Y_i^{R*} with $\mathcal{Y}^* = \{0, \ldots, 10\}$.

Hiring Managers ($N = 500$) were shown a randomly-selected Worker's gender G_i and a Recruiter's wage offer to them.[43] Formally, each Hiring Manager j 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 methodology as before. We denote the Hiring Manager's action (wage offer) as A_{ij}^H, with $\mathcal{A}$ as defined above.

6.1.1 Results

We measure discrimination with respect to Worker qualification $Y_i^{R0} = Y_i^{H0} = S_i^R$ (Task B performance), with $\mathcal{Y}^0 = \{2, 3, 4, 5, 6\}$. This focuses our measure on disparities arising between Workers who enter the hiring market with the same initial productivity signal.

Workers: There were no significant gender differences in Worker performance on either Task. On average, Workers completed 3.47 questions correctly on Task A and 3.63 questions correctly on Task B. Regressing overall performance (the sum of performance on both tasks) on a gender dummy (Male=1) yields a 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; $p = 0.34$) on a gender dummy. Performance on Task B was predictive of performance on Task A. Regressing the latter on the former yields a coefficient of 0.37 ($p < 0.01$). 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 ($p = 0.54$).

Recruiters: Given qualification $Y_i^{R0} = S_i^R$, any discrimination by Recruiters is direct. We can rule out accurate statistical discrimination as a driver of such direct discrimination, as the signal is equally informative for both men and women. Any direct discrimination is driven by the biased preferences or beliefs of Recruiters.

Recruiters directly discriminated against female Workers. The average offered wage was 5.23. Column 1 of Table 1 shows that male Workers were offered a 0.47 higher wage than

[40] Formally, each Recruiter j took the action of stating a wage offer A_{ij}^R, with $\mathcal{A} = \{0, \ldots, 10\}$.

[41] Here and in the next study we censor earnings at zero so that they could not be negative.

[42] Recruiters saw examples of the mechanism and passed comprehension checks before making wage offers.

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

Table 1. Signal Inflation: Direct Discrimination in Recruiter Wage Offers

	(1)	(2)	(3)
Gender (1=Male, 0=Female)	0.47***	0.47***	0.49***
	(0.12)	(0.12)	(0.12)
Signal S_i^R		0.49***	0.52***
		(0.09)	(0.09)
Constant	4.99***	3.04***	5.71***
	(0.14)	(0.36)	(0.60)
Recruiter Demographic Controls	N	N	Y

Notes: This table reports coefficients from regressing Recruiter wage offers on Worker gender and the Worker's Task A performance. Columns 3 controls for Recruiter characteristics: age, gender, employment status, an indicator for the Recruiter being white, and an indicator for being college-educated. The sample includes 201 Recruiters, each evaluating two Workers ($N = 402$). Standard errors, clustered at the Worker level, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

female workers, on average ($p < 0.01$).[44] 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 ($p < 0.01$).[45] Column 3 shows we get similar results controlling for Recruiter characteristics (gender, age, race, education, and employment status).[46]

Figure 2 illustrates the direct discrimination by Recruiters, plotting average wage offers by Worker gender and Task A performance. Recruiter discrimination is similar across different performance signals. While higher signals leads to higher wage offers, there is a persistent wage gap between male and female Workers.

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 $S_i^H = s$) or systemic (i.e. stemming from male and female workers with the same Recruiter signal realization $S_i^R = s$ who then receive different Recruiter wage offers on average).

Hiring Managers discriminated against female Workers. The average Hiring Manager wage offer was 5.66, similar to the average Recruiter wage offer. Column 1 of Table 2 shows that male Workers were offered a 0.92 higher wage than female workers, on average

[44]Since each Recruiter made offers to multiple Workers, standard errors are clustered at the individual level for all analyses that follow.

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

[46]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).

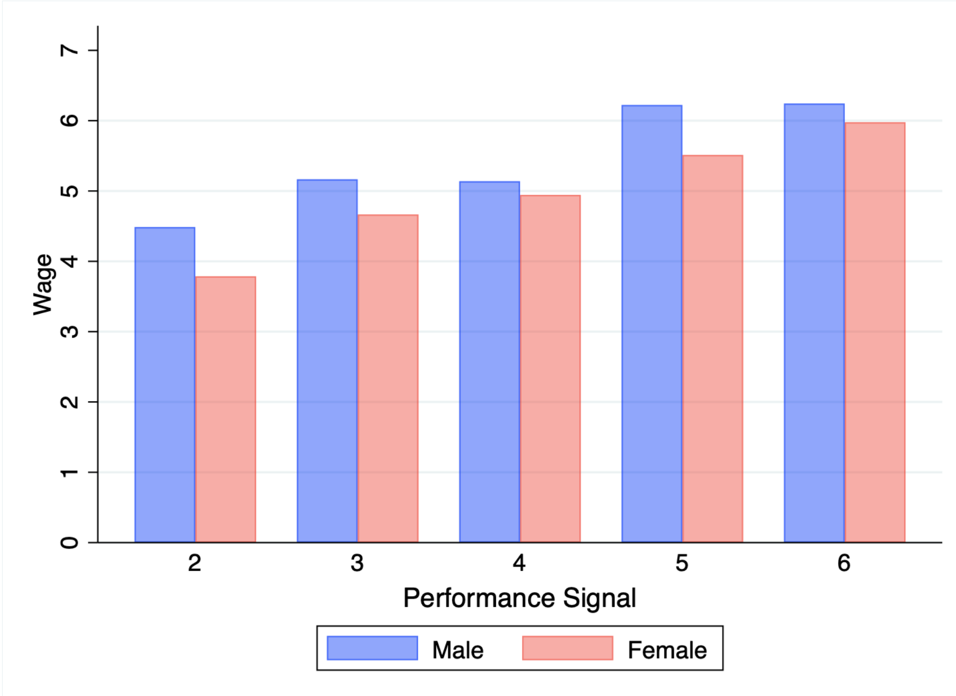

FIGURE 2. Signal Inflation: Recruiter Wage Offers by Worker Gender and Signal

($p < 0.01$). This disparity captures average total discrimination by Hiring Managers, and corresponds to roughly 0.45 standard deviations of Hiring Manager wage offers.

Hiring Manager discrimination, however, is mostly systemic. This result is suggested by Column 2 of Table 2: adding a control for the Hiring Manager signal (i.e. the Recruiter wage offer) to the regression makes the effect of gender small and insignificant. Hiring Managers therefore appear to offer similar wages to male and female Workers with the same Hiring Manager signal realization, on average. Columns 3 and 4 add Managers' demographic variables, again yielding similar results.

We now proceed to quantify systemic discrimination using the decompositions in Section 5.1. We first estimate total Hiring Manager discrimination $\Delta(y)$ by comparing male and female wage offers for each Task A performance level $y \in \{2, 3, 4, 5, 6\}$. We then estimate Hiring Managers' average direct discrimination against female workers with a given Task A performance, $E[\tau_i \mid G_i = m, Y_i^0 = y]$, by averaging gender disparities across each Hiring Manager signal realization according to the distribution each Task B 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 the measure of systemic discrimination. 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 3 confirms that most Hiring Manager discrimination is systemic. Estimated total

Table 2. Signal Inflation: Total Discrimination in Hiring Manager Wage Offers

	(1)	(2)	(3)	(4)
Gender (1=Male, 0=Female)	0.92***	-0.09	0.95***	-0.09
	(0.19)	(0.13)	(0.20)	(0.13)
Signal S_i^H		0.72***		0.72***
		(0.03)		(0.03)
Constant	5.18***	1.78***	5.36***	1.76***
	(0.14)	(0.15)	(0.42)	(0.30)
Hiring Manager Demographic Controls	N	N	Y	Y

Notes: This table reports coefficients from regressing Hiring Manager wage offers on self-reported Worker gender and the Worker's Recruiter wage offer. Columns 3 and 4 control for Manager characteristics: age, gender, employment status, an indicator for the Manager being white, and an indicator for being college-educated. The sample includes 506 Hiring Managers, each evaluating one Worker ($N = 506$). Robust standard errors are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

discrimination against female Workers ranges from 0.47 to 2.01 for Task B performance levels 2 through 6. Estimated average direct discrimination is smaller (and sometimes negative) in each decomposition. For example, Column 1 shows that total discrimination is 1.00 for Workers with a Task B performance level of 2. Estimates of systemic discrimination for this performance level range from 1.10 to 1.25, while estimated average direct discrimination at this performance level ranges from -0.25 to -0.10. At the highest level of Task B performance in the table, total discrimination is 0.33 with systemic discrimination ranging from 0.20 to 0.22 and average direct discrimination ranging from 0.10 to 0.13.

The smaller levels of direct discrimination observed in the case of Hiring Managers versus Recruiters likely stems from their beliefs about the informativeness of the provided signal. As discussed in Bohren et al. (2019), belief-based discrimination is predicted to decrease as the perceived precision of the signal increases. Since the signal for males is inflated relative to the signal for females, accurate statistical discrimination would favor female workers to undo this inflation. Given that the direct discrimination mostly favors males, it must stem from biased beliefs or preferences (Bohren et al. 2020).

6.2 Screening

Our second experiment examined the role of screening in generating systemic discrimination. The study used the same group of Workers as the first experiment. The procedures for the two sets of managers were similar except for a few notable differences. First, neither manager was given an initial endowment. As in the first study, a new group of Recruiters ($N = 200$) were shown the Task A performance of two Workers and the Workers' gender. The main difference is that, rather than stating their highest willingness to pay, Recruiters here were

Table 3. Signal Inflation: Total, Direct, and Systemic Discrimination in Manager Wages

	(1)	(2)	(3)	(4)	(5)
	Worker Performance Level Y_i^{H0}				
	2	3	4	5	6
Total Discrimination	1.00	1.39	0.47	2.01	0.33
Equation (4)					
Average Direct Discrimination	-0.10	0.30	0.11	0.51	0.10
Systemic Discrimination	1.10	1.09	0.36	1.50	0.23
Equation (5)					
Average Direct Discrimination	-0.25	-0.17	0.29	-0.08	0.13
Systemic Discrimination	1.25	1.56	0.17	2.08	0.20
Equation (6)					
Average Direct Discrimination	-0.18	0.07	0.20	0.22	0.12
Systemic Discrimination	1.18	1.33	0.27	1.79	0.22

Notes: This table reports estimates of total discrimination, average direct discrimination, and systemic discrimination in Hiring Manager wage offers across different levels of Worker productivity. Total discrimination is measured by the average difference in wage offers among male vs. female Workers with a given Task A score. Average direct and systemic discrimination are measured by equations Equations (4) to (6), as described in the text. The sample includes 506 Hiring Managers, each evaluating one Worker ($N = 506$).

asked to select which of the two Workers they would like to hire. They were then paid 1 dollar for each question the hired Worker answered correctly on Task B, above 5. Recruiter j's action rule is thus $A_{ij}^R \in \{0, 1\}$ and their payoff Y_i^{R*} is based on Task B performance, with $\mathcal{Y}^* = \{0, \ldots, 10\}$.

Hiring Managers ($N = 500$) saw one Worker's profile who had been evaluated by a Recruiter, along with information about their gender. They were shown information on the Worker's Task A performance but *only* if the Recruiter had chosen to hire them; Managers saw no performance information if the Worker had not been hired. Then, Hiring Managers made a binary decision of whether or not to hire the Worker. If hired, the Manager received a bonus corresponding to the Worker's Task B performance; if not hired, the 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 ($A_{ij}^R = 1$). If the Worker was not hired ($A_{ij}^R = 0$), the Hiring Manager observed no signal ($S^H = \emptyset$). In either case, Hiring Managers saw Worker gender G_i. The Manager's action $A_{ij}^H \in \{0, 1\}$ corresponds to her hiring the Worker. As before, the relevant measure of Worker productivity for Hiring Managers is Task B performance, $Y_i^{H*} \in \{0, \ldots, 10\}$.

6.2.1 Results

We measure systemic and total discrimination with respect to Task B performance. Total discrimination for recruiters at signal realization $S_i^R = s^R$ is again equal to the level of direct discrimination at this qualification level. We first discuss Recruiter direct discrimination before discussing direct and systemic discrimination in Hiring Manager actions.

Recruiters: Recruiter hiring actions exhibited direct discrimination 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%.[47] Given the lack of gender-based performance differences, as reported in Section 6.1.1, 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: Differential hiring rates in A_i^H reflect (total) discrimination, which can be direct (i.e. holding S_i^H fixed) or systemic. As before, systemic discrimination stems from the dependence of the Hiring Manager's signal on the Recruiter's action. Here, however, Recruiter actions affect signal *informativeness*—whether the Hiring Manager sees an (uninflated) productivity signal. Per Section 4.5.1 we expect the differences in signal quality across groups to lead to heterogeneity in the systemic discrimination faced by Workers with different qualification (Task B performance) levels.

We find significant discrimination against female workers by Hiring Managers. 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 22%. This average effect masks important heterogeneity. Figure 3 shows hiring rates by Worker gender and their Task A performance. While the discrepancy in hiring rates is relatively small for low performance levels, it increases substantially at high performance: the gender gap in hiring rates increases from 5 to 27 percentage points as we move from the lowest to the highest Task A performance levels.

The heterogeneity in total discrimination is due in part to gender differences in the availability of productivity signals. Because of direct discrimination by Recruiters, Managers are 27% less likely to see Task A performance information from a high-performing woman than from an equally qualified man. Hiring Managers are substantially more likely to hire a high-performing Worker when they have access to performance information versus the case where they do not (53% vs. 12%, $p < 0.01$). Since Managers are less likely to learn this information about women, they are less likely to hire them. In this way, systemic discrimination from screening hurts high-performing women the most.

As before we estimate systemic discrimination in Hiring Manager actions using the de-

[47] Standard errors are clustered at the individual level. Since each Recruiter had to make an offer to one of the Workers, we do not include further controls when examining hiring rates.

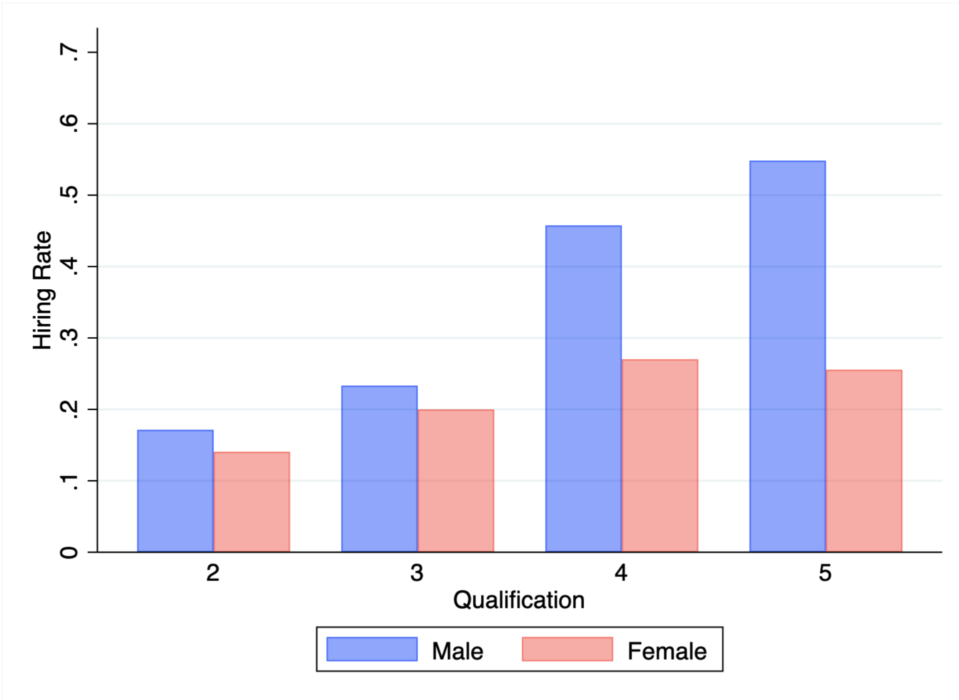

FIGURE 3. Screening: Manager Hiring Rates by Worker Gender and Qualification

compositions in Section 5.1. We first estimate total Hiring Manager discrimination $\Delta(y^0)$ by comparing male and female hiring rates based on Task B performance. We then estimate the average direct Hiring Manager discrimination $E[\tau_i \mid G_i = m, Y_i^{H0} = y^0]$ faced by male Workers with a given Task B performance y^0 by computing and averaging gender disparities that condition on the observed y^0 (or nothing). Per Equation (4), subtracting this estimate of direct discrimination from the estimate of total discrimination yields an estimate of the measure of systemic discrimination. We similarly decompose total discrimination into the alternative measures of direct and systemic components in equations Equations (5) and (6). Again, see Appendix A for details on these three calculations.

Table 4 confirms the heterogeneity in systemic discrimination faced by women with different productivity levels. At the two lower levels of Task B performance, systemic discrimination is estimated to be slightly negative or zero. However, we estimate positive systemic discrimination at the two higher performance levels, ranging from 2 to 3 percentage points when $Y_i^{H0} = 4$ and from 7 to 11 percentage points when $Y_i^{H0} = 5$. Thus, only looking at direct discrimination would miss up to 40 percent of total discrimination in our setting. Interestingly, estimated direct discrimination also rises with Worker productivity; the heterogeneity in Figure 3 comes from both types of discrimination in Hiring Manager actions.

In summary, our first two empirical investigations 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 correspon-

Table 4. Screening: Total, Direct, and Systemic Discrimination in Hiring Manager Actions

	(1)	(2)	(3)	(4)
	Task B Performance Y_i^{H0}			
	2	3	4	5
Total Discrimination	0.05	0.09	0.16	0.27
Equation (4)				
Average Direct Discrimination	0.06	0.09	0.14	0.20
Systemic Discrimination	-0.01	0.00	0.02	0.07
Equation (5)				
Average Direct Discrimination	0.07	0.10	0.13	0.16
Systemic Discrimination	-0.02	0.00	0.04	0.11
Equation (6)				
Average Direct Discrimination	0.06	0.09	0.13	0.18
Systemic Discrimination	-0.01	0.00	0.03	0.09

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 B. Total discrimination is measured by the average difference in hiring rates among male vs. female Workers with a given Task A score. Average direct and systemic discrimination are measured by equations Equations (4) to (6), as described in the text. The sample includes 501 Hiring Managers, each evaluating one Worker ($N = 501$).

dence 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 much smaller than total discrimination in the first study and misses important heterogeneity in total discrimination in the second study. The results also underscore the pitfalls of conditioning on observables which 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. 2020) may still allow discrimination to persist through systemic factors.

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

(A) No 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

(B) Work Experience

Figure 4. Example Resumes

7 Iterated Audit

Our lab-in-the-field experiment follows the IA design developed in Section 5. We used the *Centiment.co* platform to recruit hiring managers ($N = 208$) with experience in evaluating applicants to entry level jobs.[48] In a factorial ratings design that largely follows 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.[49] This 1-10 hiring likelihood score is the main dependent variables in our analyses. Decisions were incentivized using a similar methodology as Kessler et al. (2019), where 1 in 20 managers would receive a resume from an actual applicant that best matched their highest-rated fictitious applicant from the study.

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 signals of group identity—featuring distinctively Black or white names—and previous work experience. Example resumes are depicted in Figure 4; Figure 4b features relevant work experience while Figure 4a does not. The list of names and proportion of resumes with previous work experience were both taken from the results of a previously run audit study by Pager (2003), who examined the effect of group identity on hiring prospects for entry level jobs. There, Black applicants were significantly less likely to proceed through the application process than white applicants with the same qualifications. The two sets of resumes in our study, which we refer to as white-endogenous (set A) and Black-endogenous (set C), feature entries for previous work experience that match the group-based differences documented in Pager (2003). Specifically, 32% of white-endogenous

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

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

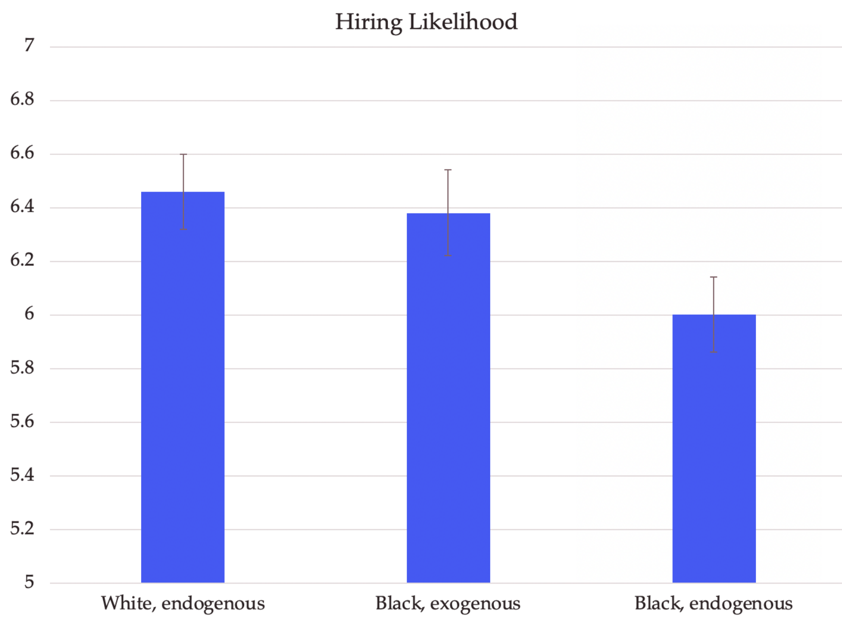

FIGURE 6. Hiring Likelihoods in Iterated Audit

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. Evaluators were told about the results of the Pager (2003) study and its connection to the task at hand prior to making their evaluations.

7.1 Results

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

Figure 6 presents results on the average hiring likelihoods by resume type. Table 5 presents these results in regression form. We find significant total discrimination. Column 1 compares white-endogenous (A) to Black-endogenous (C) resumes and reveals a gap of 0.5 in the likelihood of getting hired. This corresponds to 21% of one standard deviation in hiring likelihood. Adding hiring manager demographic controls does not diminish the gap (Column 2). 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 3 shows that the gap in hiring likelihood decreases substantially and becomes insignificant when these controls are added.

Comparing white-endogenous (A) to Black-exogenous (B) in Column 4 reveals that the

Table 5. Iterated Audit

	(1)	(2)	(3)	(4)
Race (1=Black, 0=white)	-0.46**	-0.47**	-0.30	-0.04
	(0.20)	(0.20)	(0.21)	(0.20)
Prior Experience			1.26***	
			(0.21)	
Constant	6.46***	5.96***	5.61***	6.14***
	(0.17)	(0.74)	(0.757)	(0.80)
Demographic Controls	N	Y	Y	Y
Exogenous-Black Resumes (B)	N	N	N	Y

Notes: This table reports coefficients from regressing hiring likelihood on resume type. Columns 1 compares types A to C; Column 2 controls for Manager characteristics: age, gender, an indicator for the Manager being white, and job duration. Column 3 compares types A to C while controlling for prior work experience and Manger characteristics. Column 4 compares types A to B and controls for Manager characteristics. The sample includes 208 Hiring Managers, each evaluating four different resumes. Manager-clustered standard errors are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

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 (Column 4 of Table 5), systemic discrimination still led to substantial differences in hiring likelihoods. In light of prior work showing the effectiveness of information in reducing direct discrimination (Bohren et al. 2020), these findings highlight the difficulty of mitigating total discrimination when it is caused by systemic factors.

8 Conclusion

Vast literatures in social and computer science emphasize the importance of systemic factors in driving group-based disparities, yet economic analyses largely focus on direct discrimination by individuals. This paper seeks 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.

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. 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%. 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.[50] 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.

[50]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.

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.

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

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 (2020): "Inaccurate Statistical Discrimination: An Identification Problem," *NBER Working Paper No. 25935.*

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.

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.

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.

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*.

Glover, D., A. Pallais, and W. Pariente (2017): "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*.

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

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.

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.

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.

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 $y \in \{2,3,4,5,6\}$ in the first experiment and $y \in \{2,3,4,5\}$ in the second experiment, total discrimination is given by

$$\Delta(y) = E[A^H_{i,j(i)} \mid G_i = m, Y_i^{H0} = y] - E[A^H_{i,j(i)} \mid G_i = f, Y_i^{H0} = y]$$

where $j(i)$ denotes the Hiring Manager of Worker i, $A^H_{ij(i)}$ is the Hiring Manager action for Worker i, G_i is Worker i's self-reported gender (either male m or female f), and Y_i^{H0} is Worker i's qualification (Task A performance). We estimate total discrimination by the corresponding sample average differences, $\widehat{\Delta}(y)$.

Expected direct discrimination at Hiring Manager signal realization $s^H \in \mathcal{S}^H$, where $\mathcal{S}^H = \{1,\ldots,10\}$ 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

$$\tau(s) = E[A^H_{i,j(i)} \mid G_i = m, S_i^H = s^H] - E[A^H_{i,j(i)} \mid G_i = f, S_i^H = s^H],$$

with $\tau_i \equiv \tau(S_i^H)$ giving the expected direct discrimination faced by each worker i. The corresponding sample average differences $\hat{\tau}(s)$ yield estimates $\hat{\tau}_i = \hat{\tau}(S_i^H)$. For the first term of Eq. (4) we then compute average direct discrimination as

$$\hat{E}[\tau_i \mid G_i = m, Y_i^{H0} = y] = \frac{1}{N_{m,y}} \sum_{i:G_i=m,Y_i^0=y} \hat{\tau}_i,$$

for each y, where $N_{g,y}$ gives the number of Workers with $G_i = m$ and $Y_i^{H0} = y$. This gives our estimates of average direct discrimination for Equation (4) in Table 3. Estimates of systemic discrimination are then given by

$$\hat{\delta}(f,y) = \widehat{\Delta}(y) - \hat{E}[\tau_i \mid G_i = m, Y_i^{H0} = y]$$

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

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

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

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 $\mu_m > \mu_f$. 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.[51] 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 $(\mu_m - \mu_f)/2 > 0$. There is no systemic discrimination, because conditional on productivity the signal process is the same for group-m and group-f 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, $E[A_2(g, S_{i2}) \mid G_i = m, Y_i^*] = (\mu_m + 2Y_i^*)/3 > (\mu_f + 2Y_i^*)/3 = E[A_2(g, S_{i2}) \mid G_i = f, Y_i^*]$, so systemic discrimination is given by $(\mu_m - \mu_f)/3 > 0$. Similarly, systemic discrimination persists in subsequent periods: in period t, $E[A_t(g, S_{it})|G_i = m, Y_i^*] = (\mu_m + tY^*)/(t+1) > (\mu_f + tY^*)/(t+1) = E[A_t(g, S_{it})|G_i = f, Y_i^*]$, yielding systemic discrimination of $(\mu_m - \mu_f)/(t+1)$.[52] Intuitively, the initial accurate statistical discrimination

[51] 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.

[52] 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.

B.2 Institutional Systemic without Institutional Direct

This example shows how individual direct discrimination can lead to institutional systemic discrimination in another market, despite not aggregating to institutional direct discrimination. Consider the setting from Section 2. Suppose Recruiters observe a productivity signal which has a distribution that does not differ by worker group, $F_s^R(s^R|y,m) = F_s^R(s^R|y,f) = F_s^R(s^R|y)$. Let $f_s^R(s^R|y)$ denote the associated density. There are two types of recruiters in the firm, 1 and 2. Recruiter type 1 has action rules $A^{R,1}(f, s^R) = s^R$ and $A^{R,1}(m, s^R) = s^R + 2$, resulting in direct discrimination against group f. Recruiter type 2 has action rules $A^{R,1}(f, s^R) = s^R + 1$ and $A^{R,1}(m, s^R) = s^R$, resulting in direct discrimination against group m. Suppose the share of type-1 and type-2 recruiters is 1/3 and 2/3, respectively. Then there is no institutional direct discrimination among recruiters: $\alpha_1(g, s^R) = s^R + 2/3$ for each $g \in \{m, f\}$. But the distribution of hiring manager signals (i.e. recruiter actions) does depend on g: the corresponding density is given by $f_s^H(s^H|y,f) = \frac{1}{3}f_s^R(s^H|y) + \frac{2}{3}f_s^R(s^H - 1|y)$ and $f_s^H(s^H|y,m) = \frac{1}{3}f_s^R(s^H - 2|y) + \frac{2}{3}f_s^R(s^H|y)$. Thus, there is institutional systemic discrimination.

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 4.5.1).