Systemic Discrimination: Theory and Measurement[*]

J. Aislinn Bohren[†] Peter Hull[‡] Alex Imas[§]

March 7, 2022

Economics tends to define and measure discrimination as disparities stemming from the direct (causal) effects of protected group membership. But work in other fields notes that such measures are incomplete, as they can miss important systemic (i.e. 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 both direct and systemic forms of discrimination. We define systemic discrimination as emerging from group-based differences in non-group characteristics, conditional on a measure of individual qualification. We formalize sources of systemic discrimination as disparities in signaling technologies and opportunities for skill development. Notably, standard tools for measuring direct discrimination, such as audit or correspondence studies, cannot detect systemic discrimination. We propose a measure of systemic discrimination based on a novel decomposition of total discrimination—disparities that condition on underlying qualification—into direct and systemic components. This decomposition highlights the type of data needed to measure systemic discrimination and guides identification strategies in both observational and (quasi-)experimental data. We illustrate these tools in two hiring experiments. 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 David Arnold, Leonardo Bursztyn, Diag Davenport, Will Dobbie, Sam Hirshman, Matthew Knepper, Matt Lowe, Heather Sarsons, Damián Vergara, and seminar participants at Brown for helpful comments. Christy Kang 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 studied 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 be a product of 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 (2021b); (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 more concrete 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—explicitly differential group-based treatment—and *systemic* discrimination: group-based differences in non-group characteristics that indirectly lead to unequal treatment. Both forms contribute to *total* discrimination: treatment disparities among equally qualified individuals. Depending on the selected measure of qualification, this framework can be used to study different sources of systemic discrimination. 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 are 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 biases in preferences and beliefs. Institutional direct discrimination is generated through

[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.

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.

We formalize two conceptually distinct 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 develop a new measure of systemic discrimination that leverages a novel decomposition of total discrimination into direct and systemic components, building on the classic decompositions of Kitagawa (1955), Oaxaca (1973), and Blinder (1973). The direct discrimination component can be identified through standard methods in economics which condition on or randomize over all relevant non-group characteristics. Total discrimination is identified when an individual's qualification for treatment is observed and can be conditioned on. More generally, this component can be measured when the joint distribution of qualification and group membership is identified through, for example, quasi-experimental methods (Arnold, Dobbie, and Hull 2021a; Arnold et al. 2021b). Our measure of systemic discrimination is then given by the residual of the identified decomposition. The decomposition thus delineates the type of data needed to measure systemic discrimination and can be used to guide identification strategies in observational and (quasi-)experimental data.

We illustrate how our decomposition can be used in practice using two pre-registered hiring experiments. In each experiment, participants were randomized into one of three roles: Worker, Recruiter, and Hiring Manager. Workers completed two Tasks, A and B, each consisting of questions on different subjects. Recruiters evaluated Workers and took actions based on information about their performance on Task A and the Workers' self-reported gender identity. Recruiters were paid on the basis of the Workers' performance on Task B and their actions. Hiring Managers also evaluated Workers and took actions after observing Worker gender and a performance signal. Critically, however, the latter non-group signals were determined endogenously through Recruiter actions, allowing direct discrimination by Recruiters to generate systemic discrimination in Hiring Manager actions through the signaling technology.

This unique experimental design allows us to quantify two different forms of systemic discrimination via our decomposition. In the first "signal inflation" experiment, the actions corresponded to wage offers. Hiring Managers observed Recruiter wage offers as a performance signal along with Worker gender. This experiment allows us to examine systemic discrimination generated by disparities in the performance signals themselves. In the second "screening" experiment, the actions were hiring decisions. Hiring Managers observed Worker performance signals *only if* the Worker had been hired by a Recruiter. We use this experiment to study systemic discrimination arising from the disparate availability of objective performance signals and group-based differences in information precision.

Both studies revealed significant direct and systemic discrimination. In the first study, Recruiters made lower wage offers to female Workers than male Workers with similar performance signals. Because we did not find gender differences in actual Worker performance on either task, these disparities represent direct discrimination—either due to Recruiter preferences or inaccurate beliefs (Bordalo, Coffman, Gennaioli, and Shleifer 2019; Bohren, Imas, and Rosenberg 2019). We also found substantial (total) discrimination in wage offers made by Hiring Managers. However, our decomposition shows that the vast majority of this discrimination was systemic and not direct. Holding Hiring Managers' information about Worker qualification (i.e. Recruiter wage offers) fixed, female Workers received only slightly lower wage offers than male workers. But direct discrimination by Recruiters caused the signals that Hiring Managers saw about male Workers to be inflated relative to the signals of female workers, which led them to make higher wage offers to male Workers. Standard discrimination measures that condition on non-group characteristics would miss this systemic discrimination, and thus the majority of total discrimination in this setting. Our findings also illustrate a critical implication of our framework for the long-run effects of inaccurate beliefs or biased stereotypes: initial biases can drive persistent disparities through systemic channels, even if direct discrimination is mitigated.

Our second study also revealed substantial direct discrimination in Recruiters' hiring decisions: holding performance signals constant, male Workers were more likely to be hired than female Workers. This discrimination in the screening of Workers contributed to the total discrimination in Hiring Manager actions, particularly in the case of high performing women. Hiring Managers could only learn about a Worker's performance if they were hired by the Recruiter; since well-qualified women were less likely to be hired by Recruiters than well-qualified men, Hiring Managers were less likely to learn about their high performance. Systemic differences in signal informativeness thus resulted in substantially lower hiring rates for high-performing women than similarly qualified men. These findings illustrates the scope for important heterogeneity in how screening decisions can impact total discrimination.

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 6 presents our empirical investigation. Section 7 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.[4] 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).[5]

[4] 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.

[5] 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 "color-

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)*.[6] 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.[7]

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 Rodriguez, 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.[8] 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

blind" treatment because they were aware systemic discrimination could preserve the status quo.

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

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

[8]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).

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. 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).[9] 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 thus 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.[10]

[9] 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, which we formalize below.

[10] 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.

2 A Motivating Example

We begin our analysis with a simple motivating 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 $G = f$. 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.[11]

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.[12] 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 systematic disparities in actions A^H that stem indirectly from the dependence of non-group signal S^H on

[11] Bias, either in the form of taste-based discrimination or inaccurate statistical discrimination, 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.

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

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.

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. Thus 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.[13]

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

[13]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.

$G_i \in \{m, f\}$ 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.[14] 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 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

[14] 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.

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.[15]

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:

[15]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.

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 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).[16] 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.[17] 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.[18]

[16]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.

[17]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).

[18]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.

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

3.3 The Choice of Y_i^0.

Both systemic and total discrimination are defined with respect to the chosen measure of worker qualification Y_i^0, and are thus inherently tied to the researcher's choice of this reference point. At one extreme, when worker qualification is set equal to non-group characteristics observed by the manager ($Y_i^0 = S_i$), total discrimination is narrowly defined as any treatment disparities that remain when holding fixed the relevant 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.[19]

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

[19] 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.

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

[20] 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).

[21] 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.

[22] Arnold et al. (2021b), 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 et al. 2021a).

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.

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).[23] 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

[23] 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.

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 (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 thus 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.[24] Group f may also have less informative productivity signals because they had less previous 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.[25] 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.[26]

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

[24]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.

[25]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.

[26]Analogous to how direct discrimination can arise from omitted payoff bias (see Footnote 23), 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.

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.[27]

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 dis-

[27]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.

crimination 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.[28] 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

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

of 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).

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 thus 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.[29] Since $\eta_f \neq \eta_m$, this expression is nonzero unless $y = t\frac{1+\eta_g}{\eta_g}$. Thus, 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

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

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 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 thus 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 $E[A_1(g, S_i)|Y_i^0 = y, G_i = m] - E[A_1(g, S_i)|Y_i^0 = y, G_i = f] = 1 - \gamma > 0$.[30] 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 dis-

[30] 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 $g = f$.

crimination 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: $A^{H,m}(f,s) = s$, $A^{H,m}(m,s) = s+1$, $A^{H,f}(f,s) = s+1$ 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 Identification and Decomposition

We now discuss challenges and solutions in measuring direct, systemic and total discrimination. We first discuss existing approaches to identifying direct discrimination and certain kinds of total discrimination. We then propose a new strategy of measuring systemic discrimination by decomposing total discrimination into direct and systemic components. We bring this decomposition to data in Section 6.

Core identification challenges stem from the fact that the econometrician may not observe all relevant worker characteristics, (G_i, Y_i^*, S_i, Y_i^0). Our baseline analysis supposes the econometrician observes worker groups G_i, manager actions A_{ij}, and an ex post measure of output Y_{ij} that is determined by manager actions and worker productivity Y_i^*. For example, when $A_{ij} \in \{0, 1\}$ indicates a hiring decision, output may be given by $Y_{ij} = A_{ij} Y_i^* \in \{0, Y_i^*\}$. Workers who are not hired generate no output, while hired workers translate their productivity to output. In some of the discussion we also consider the observability of the non-group signal S_i and qualification measure Y_i^0.

For simplicity we assume $\mathcal{A} \subset \mathbb{R}$ throughout this section and focus on measures of discrimination corresponding to mean differences by group.[31]

5.1 Measuring Direct Discrimination

A measure of direct discrimination by manager j at signal realization $s \in \mathcal{S}$ is given by

$$\tau_j(s) \equiv A_j(m, s) - A_j(f, s), \tag{1}$$

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.

The core challenge with measuring direct discrimination is the unobservability of manager signals: if S_i is observed, direct discrimination is identified by a simple conditional disparity in average actions. Specifically, $E[A_{ij} \mid G_i = m, S_i = s] - E[A_{ij} \mid G_i = f, S_i = s]$ identifies the individual direct discrimination exhibited by manager j at each $s \in \mathcal{S}$. Signals may be observed when managers follow a known algorithmic action rule, or when manager information sets are otherwise fully under the econometrician's control (as in certain correspondence studies). In general, however, managers are likely to act on a range of subjective or otherwise hard-to-measure signals which are not easily conditioned on. Feasible conditional disparities are then likely to suffer from omitted variables bias (OVB), relative to direct discrimination.

The typical solution to this OVB challenge is a correspondence study where the manager's perception of the worker's group is randomized to be independent of any other information

[31] Our analysis of means generalizes to other distributional features of A_{ij}, such as mean disparities in the indicators $\mathbb{1}[A_{ij} \le a]$ for $a \in \mathcal{A}$.

that might enter S_i. For example, the econometrician may randomly vary a worker's self-reported gender, holding other characteristics in her employment application fixed.[32] To formalize this solution, let $\tilde{G}_i$ denote the managers' perception of G_i after such randomization, with $\tilde{G}_i \perp\!\!\!\perp (G_i, Y_i^*, S_i, Y_i^0)$ and now $A_{ij} = A_j(\tilde{G}_i, S_i)$.[33] Then, even when S_i is not fully observed, average direct discrimination is identified by simple $\tilde{G}_i$-based disparities. For example, the unconditional difference in means

$$\begin{aligned} E[A_{ij} \mid \tilde{G}_i = m] - E[A_{ij} \mid \tilde{G}_i = f] &= E[A_j(m, S_i) \mid \tilde{G}_i = m] - E[A_j(f, S_i) \mid \tilde{G}_i = f] \\ &= E[A_j(m, S_i) - A_j(f, S_i)], \end{aligned}$$

identifies the average individual direct discrimination exhibited by manager j. The fact that randomization reveals direct discrimination reinforces its causal nature: $A_j(g, S_i)$ can be viewed as the potential outcome of worker i when manager j perceives her group to be g, and $E[A_j(m, S_i) - A_j(f, S_i)]$ can be viewed as the average treatment effect (ATE), of perceived group membership on manager j's actions.

Distinguishing between the different sources of direct discrimination generally requires additional (quasi-)experimental variation or outside knowledge of the action rule. For example, when S_i is not observed, a marginal outcome test which compares the productivity of workers the manager is just indifferent to treating in a particular way can distinguish accurate statistical discrimination from bias (Hull 2021; Grau and Vergara 2021). Such tests can be conducted when managers are as-good-as-randomly assigned to workers, typically under additional identifying restrictions (e.g. Arnold et al. 2018). Distinguishing between biased preferences and beliefs is generally challenging, absent direct information on (or experimental variation in) manager beliefs and signals (Bohren et al. 2020; Canay et al. 2020).

Institutional direct discrimination can be measured with a firm-level correspondence study in which perceived group membership is randomized (e.g. Kline et al. 2021). Separately measuring the direct discrimination at a firm and the direct discrimination of its managers can reveal whether the former is driven by aggregated heterogeneous behavior or uniform discriminatory decision-making, which might inform potential policy responses. For example, if institutional direct discrimination is driven by a combination of in-group favoritism and underrepresentation of a group among managers, increasing representation may be an effective policy response (see Section 4.5.3). But if institutional direct discrimination is driven by bias across all managers, effective policy responses must target bias directly.

[32] 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 15.

[33] Workers are fictional in some correspondence studies, such that the information entering S_i is also randomly generated (independently of $\tilde{G}_i$). These workers have no G_i or Y_i^* (and maybe no Y_i^0), but we still have $G_i \perp\!\!\!\perp S_i$. Average direct discrimination is then still identified when the marginal distribution of S_i in the experiment coincides with the population distribution.

In contrast, systemic discrimination cannot be measured by an experimental manipulation of perceived group membership. It arises from the statistical relationship between true group membership G_i and the non-group signal S_i, conditional on qualification Y_i^0, as well as the structure of action rules $A_j(g, s)$. Randomized membership $\tilde{G}_i$ has no relationship with S_i by design, so conventional correspondence studies cannot measure systemic discrimination. For example, systemic discrimination arising from i's membership in certain social clubs (as recorded in S_i) is obscured when club membership is randomized against or conditioned on when computing racial or gender disparities in firm hiring rates.

5.2 Measuring Total Discrimination

A measure of total discrimination by manager j at qualification $y^0 \in \mathcal{Y}^0$ is given by

$$\mathit{\Delta}_j(y^0) \equiv E[A_j(G_i, S_i) \mid G_i = m, Y_i^0 = y^0] - E[A_j(G_i, S_i) \mid G_i = f, Y_i^0 = y^0], \tag{2}$$

with an analogous definition of $\mathit{\Delta}(y^0)$ with respect to $\alpha(g, s)$ for firm-level (institutional) total discrimination. A finding of $\mathit{\Delta}(y^0) > 0$ 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.

The core challenge with measuring total discrimination is the unobservability of worker qualification: if Y_i^0 is observed, total discrimination is identified by a simple conditional disparity in average actions. Specifically, $E[A_{ij} \mid G_i = m, Y_i^0 = y^0] - E[A_{ij} \mid G_i = f, Y_i^0 = y^0]$ identifies the individual (total) discrimination exhibited by manager j at each $y^0 \in \mathcal{Y}^0$. Qualification may be observed when it is chosen to be a simple predetermined characteristic, such as a worker's educational attainment prior to joining the labor market. In the extreme case of $Y_i^0 = 0$, total discrimination is identified simply by the unconditional disparity in A_{ij}. In general, however, worker qualification is likely to be at best selectively observed. 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 hiring, observed output $Y_{ij} = A_{ij} Y_i^*$ gives a selective measure of qualification: workers who are hired ($A_{ij} = 1$) reveal their qualification, 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.

Arnold et al. (2021a,b) develop quasi-experimental solutions to the challenge of selectively observable qualification, with examinations of disparate impact in pretrial release decisions leveraging the as-good-as-random assignment of pretrial judges. To translate their approach to the employment setting, let $Y_i^0 = Y_i^*$ be a binary measure of worker productivity and let A_{ij} be a binary hiring decision with $Y_{ij} = A_{ij} Y_i^*$. When managers are as-good-as-randomly assigned to workers, identification of the group-specific qualification means $E[Y_i^* \mid G_i]$ is sufficient to measure disparate impact (i.e. total discrimination) in hiring decisions at either the individual or institutional level.[34] Arnold et al. (2021b) further show that these key moments can be estimated by leveraging the quasi-random assignment of managers with different hiring rates. This approach effectively "selection-corrects" the observed worker output with statistical extrapolations across the quasi-randomly assigned managers, in an instrumental variables approach similar to Heckman (1990).

Quasi-experimental assignment can thus be used to identify total (either individual or institutional) discrimination when Y_i^0 is selectively observed. In the $Y_i^0 = Y_i^*$ case the approach would be a direct application of Arnold et al. (2021a,b), leveraging the assignment of managers in the task at hand. More generally, the quasi-experimental assignment of managers and firms in different tasks can be used to address the selection challenge for other qualification measures. For example when Y_i^0 is a measure of the qualification (potential output) of worker i when she first enters the labor market, initial quasi-experimental assignment to different hiring managers can be leveraged. As we discuss below, varying Y_i^0 from entry-level potential output to the payoff-relevant Y_i^* can isolate the systemic sources of discrimination that emerge from decisions at different stages of the job market.

5.3 Measuring Systemic Discrimination

A measure of systemic discrimination by manager j at qualification $y^0 \in \mathcal{Y}^0$ 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 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.

Our identification strategy for systemic discrimination leverages a statistical decomposition of total discrimination into direct and systemic components, in the same spirit as the classic decompositions of Kitagawa (1955), Oaxaca (1973), and Blinder (1973). With direct and total discrimination identified by the methods discussed above, systemic discrimination can be indirectly quantified by this decomposition. For concreteness, we derive the decomposition for the measures of institutional (i.e. firm-level) total, direct, and systemic discrimination, i.e. $\mathit{\Delta}(y^0)$, $\tau(s)$, and $\delta(g, y^0)$; the approach is analogous for the individual-

[34] Specifically, $E[A_j(G_i, S_i) \mid G_i = g, Y_i^* = 1] = \frac{E[A_{ij} Y_i^* | G_i = g]}{Pr(Y_i^* = 1 | G_i = g)} = \frac{E[Y_{ij} | G_i = g]}{E[Y_i^* | G_i = g]}$ with $Y_{ij} = A_{ij} Y_i^*$ and similarly $E[A_j(G_i, S_i) \mid G_i = g, Y_i^* = 0] = \frac{E[A_{ij}(1 - Y_i^*) | G_i = g]}{Pr(Y_i^* = 0 | G_i = g)} = \frac{E[A_{ij} - Y_{ij} | G_i = g]}{1 - E[Y_i^* | G_i = g]}$. The only moments that are not directly estimable in these formulas for individual (total) discrimination are given by $E[Y_i^* \mid G_i]$. Institutional (total) discrimination is then identified by the manager assignment shares π_j.

level measures. We have:

$$\begin{aligned}
\overbrace{\Delta(y^0)}^{\text{Total discrimination}} &= E[\alpha(m,S_i) \mid G_i = m, Y_i^0 = y^0] - E[\alpha(f,S_i) \mid G_i = f, Y_i^0 = y^0] \\
&\quad - E[\alpha(f,S_i) \mid G_i = m, Y_i^0 = y^0] + E[\alpha(f,S_i) \mid G_i = m, Y_i^0 = y^0] \\
&= E[\alpha(m,S_i) - \alpha(f,S_i) \mid G_i = m, Y_i^0 = y^0] \\
&\quad + E[\alpha(f,S_i) \mid G_i = m, Y_i^0 = y^0] - E[\alpha(f,S_i) \mid G_i = f, Y_i^0 = y^0] \\
&= \underbrace{E[\tau(S_i) \mid G_i = m, Y_i^0 = y^0]}_{\text{Average direct discrimination}} + \underbrace{\delta(f,y^0)}_{\text{Systemic discrimination}}, \qquad (4)
\end{aligned}$$

where the expectations are taken with respect to the distribution $\phi_s(s|y^0,g)$ when $G_i = g$ and $Y_i^0 = y^0$. The first line of this expression adds and subtracts $E[\alpha(f,S_i) \mid G_i = m, Y_i^0 = y^0]$ to and from the definition of $\Delta(y^0)$; the second and third lines rearrange terms and apply the definitions of $\tau(s)$ and $\delta(g,y^0)$. Equation (4) shows that total discrimination at qualification level y^0 can be written as the sum of two terms: (i) average direct discrimination across the signal space, where the average is taken with respect to the signal distribution for workers from group m with qualification level y^0, i.e. $\phi_s(s|y^0,m)$, and (ii) systemic discrimination at qualification level y^0 when the firm uses the action rule for group f.

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

$$\Delta(y^0) = E[\tau(S_i) \mid G_i = f, Y_i^0 = y^0] + \delta(m,y^0) \qquad (5)$$

by instead adding and subtracting $E[\alpha(m,S_i) \mid G_i = f, Y_i^0 = y^0]$ from the definition of $\Delta(y^0)$. Equation (5) decomposes total discrimination into the average direct discrimination with respect to the signal distribution for workers from group f, i.e. $\phi_s(s|y^0,f)$, and the systemic discrimination when the firm uses the action rule for group m, all at a given qualification level y^0. By averaging these two expressions, we obtain a third decomposition:

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

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

To see how this decomposition can be used to identify systemic discrimination, first suppose direct discrimination is constant across signal realizations: $\tau(s) = \tau$ for all $s \in \mathcal{S}$. Then the direct discrimination terms in Equations (4) to (6) are the same, and systemic discrimination is given by the difference between total and direct discrimination: $\Delta(y^0) - \tau =$

$\delta(f, y^0) = \delta(m, y^0) = \bar{\delta}(y^0)$. When τ and $\Delta(y^0)$ are identified (e.g. using the methods discussed above), systemic discrimination is also identified.

When direct discrimination varies across signal realizations, the measures of systemic discrimination in Equations (4) to (6) are identified by the difference between total discrimination and direct discrimination averaged across the signal distribution for workers from the given fixed group. Suppose first that qualification is observed, such that total discrimination can be measured simply by conditional-on-Y_i^0 disparities. The direct discrimination component of our decomposition can then be estimated with a correspondence study. Namely, with $\tilde{G}_i \perp\!\!\!\perp (G_i, S_i, Y_i^0)$, we have $E[\tau(S_i) \mid G_i = g, Y_i^0 = y^0]$ identified by $E[A_{ij} \mid \tilde{G}_i = m, G_i = g, Y_i^0 = y^0] - E[A_{ij} \mid \tilde{G}_i = f, G_i = g, Y_i^0 = y^0]$. For example the direct discrimination component of Equation (5), $E[\tau(S_i) \mid G_i = f, Y_i^0 = y^0]$, can be measured by computing the disparity in actions among group-f workers with qualification y^0 who are and are not randomized to a perceived group of m. Subtracting this conditional disparity from the total discrimination measure $\Delta(y^0)$, computed as the action disparity between group-m and group-f workers with qualification y^0, measures systemic discrimination $\delta(m, y^0)$.

When qualification is only selectively observed, a researcher can leverage a combination of the two (quasi-)experimental identification strategies. As-good-as-random manager assignment can first be leveraged to estimate the distribution of (G_i, Y_i^0) by the method of Arnold et al. (2021b), which identifies total discrimination. By applying this approach to a population of workers where perceived group membership has been randomized, and comparing the Arnold et al. (2021b) measure among group-g workers randomized to either $\tilde{G}_i = m$ or $\tilde{G}_i = f$, the appropriate average direct discrimination measure is identified.[35]

By varying Y_i^0 in this approach, the researcher can separate the two sources of systemic discrimination discussed in Section 4. When $Y_i^0 = Y_i^*$, group differences in the signaling technology are the only source of systemic discrimination. Different choices of Y_i^0 allow for systemic discrimination to also have a technological source. Therefore, through selecting multiple Y_i^0, it is possible to further decompose total discrimination into a direct component, a systemic informational component, and a systemic technological component. Identification of each component in this decomposition would follow similarly as above, and likely require additional (quasi-)experimental variation.

[35] E.g. if $Y_i^0 = Y_i^* \in \{0, 1\}$ and $Y_{ij} = A_{ij} Y_i^*$, the disparity in $E[A_j(\tilde{G}_i, S_i) \mid \tilde{G}_i = g', G_i = g, Y_i^* = 1] = \frac{E[A_{ij} Y_i^* \mid \tilde{G}_i = g', G_i = g]}{Pr(Y_i^* = 1 \mid \tilde{G}_i = g', G_i = g)} = \frac{E[Y_{ij} \mid \tilde{G}_i = g', G_i = g]}{E[Y_i^* \mid G_i = g]}$ over $g' \in \{m, f\}$ identifies $E[\tau(S_i) \mid G_i = g, Y_i^0 = 1]$, using randomization of $\tilde{G}_i$. Similarly, the disparity in $E[A_j(\tilde{G}_i, S_i) \mid \tilde{G}_i = g', G_i = g, Y_i^* = 0] = \frac{E[A_{ij}(1 - Y_i^*) \mid \tilde{G}_i = g', G_i = g]}{Pr(Y_i^* = 0 \mid \tilde{G}_i = g', G_i = g)} = \frac{E[A_{ij} - Y_{ij} \mid \tilde{G}_i = g', G_i = g]}{1 - E[Y_i^* \mid G_i = g]}$ over $g' \in \{m, f\}$ identifies $E[\tau(S_i) \mid G_i = g, Y_i^0 = 0]$. The only moments that are not directly estimable in these formulas are again given by $E[Y_i^* \mid G_i]$.

6 Empirical Demonstration

We illustrate our decomposition and measure systemic discrimination in two experiments. The first experiment illustrates systemic discrimination arising from signal inflation, as in the motivating example in Section 2. The second experiment illustrates how systemic discrimination can be heterogeneous in screening decisions, similar to the example in Section 4.5.1. In both studies, a pool of workers face evaluations from two sets of managers. Each experiment produces three variables: the true underlying productivity of the workers, initial evaluations of those workers based on group identity and a productivity signal, and second-stage evaluations based on group identity and an endogenous productivity signal that depends on the first-stage evaluations. Worker qualification is chosen so that there is no systemic discrimination in initial evaluations: total discrimination is equal to direct discrimination. This direct discrimination can lead to systemic discrimination in the second-stage evaluation, alongside any direct discrimination. For each study, we first describe the experimental design and how it maps to our theoretical framework, before describing the empirical findings.[36]

6.1 Systemic Discrimination from Signal Inflation

6.1.1 Design

The first experiment consists of three types of participants: Workers, Recruiters, and Hiring Managers. All participants were enlisted from the Prolific online crowd-sourcing platform and randomized into one of the three roles.[37]

Workers: 100 participants were selected for the role of Worker. Each Worker completed a test of their basic math, business, and history knowledge. The tests consisted of two Tasks (A and B), each with 10 randomly-selected questions from these topics. After completing the Tasks, Workers answered several demographic questions. From this survey we obtain self-reported gender identity (male or female), which we consider the group G_i. We restrict attention to workers who had Task A performance in $\{2, 3, 4, 5, 6\}$ in order to ensure enough data for each gender.

Recruiters: Around 200 participants were selected for the role of Recruiter. Each Recruiter was 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. Given the restriction on Task A performance, $\mathcal{S}^R = \{2, 3, 4, 5, 6\}$.

After viewing S_i^R and G_i, Recruiters were asked to state their willingness to pay to hire

[36] Preregistration materials can be found at https://aspredicted.org/TK7_R4J for the first experiment and at https://aspredicted.org/K3Q_RPK for the second experiment.

[37] Online crowdsourcing platforms have increasingly been used in experimental economics. See, for example, Enke and Graeber (2019), Frydman and Jin (2022), and DellaVigna and Pope (2018).

Worker i, between 0 and 10 dollars. Formally, each Recruiter j took the action of stating a wage offer A^R_{ij}, with $\mathcal{A} = \{0, \ldots, 10\}$. The Recruiter's payoff was determined by A^R_{ij} and the Worker's Task B performance. Recruiter wage offers were accepted or rejected according to the Becker-DeGroot-Marschak mechanism to ensure truthful reporting. Specifically, if the Recruiter's wage offer was larger than a randomly chosen number (uniformly distributed between 0 and 10) the Worker would be "hired." Recruiters then received a payment based on Task B performance of the hired Workers: they received 1 dollar for each question the Worker answered correctly on Task B, minus the randomly generated number.[38] If the Recruiter's wage offer was lower than the randomly generated number, the Worker would not be hired and the Recruiter's earnings would not be affected.[39] Task B performance is thus the relevant measure of Worker productivity in this first stage of the experiment, which we denote Y^{R*}_i with $\mathcal{Y}^* = \{0, \ldots, 10\}$.

Hiring Managers: Around 500 participants were selected for the role of Hiring Manager. Each Hiring Manager was shown a randomly-selected Worker's gender G_i and a Recruiter's wage offer to them.[40] Formally, each Hiring Manager j observed signal $S^H_i \equiv A^R_{ik}$ 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 with the Recruiters. We denote the Hiring Manager's action (wage offer) as A^H_{ij}, with $\mathcal{A}$ as defined above. Each Hiring Manager's payment depended on the Worker's performance on Task A (rather than Task B). Hence Task A performance is the relevant measure of Worker productivity for Hiring Managers, which we denote Y^{H*}_i with $\mathcal{Y}^{H*} = \{2, 3, 4, 5, 6\}$ since we restricted attention to workers with Task A performance in $\{2, 3, 4, 5, 6\}$.

6.1.2 Results

We measure discrimination with respect to Worker qualification $Y^{R0}_i = Y^{H0}_i = S^R_i$ (Task A 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. We first discuss the results from the Workers and the behavior of Recruiters, which yields a measure of direct discrimination in the initial decision. We then analyze the behavior of the Hiring Managers, which yields measures of both direct and systemic discrimination in the second-stage decision.

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)

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

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

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

Table 1. Signal Inflation: 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. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

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 thus 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 female workers, on average ($p < 0.01$).[41] This effect corresponds to around 0.22 standard deviations of Recruiter wage offers. Column 2 shows that Recruiters responded positively to their signal, with each additional question correctly answered in Task A leading to a higher wage offer of 0.49 on average ($p < 0.01$).[42] Column 3 shows we get similar results controlling for Recruiter characteristics (gender, age, race, education, and employment status).[43]

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

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

[43]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

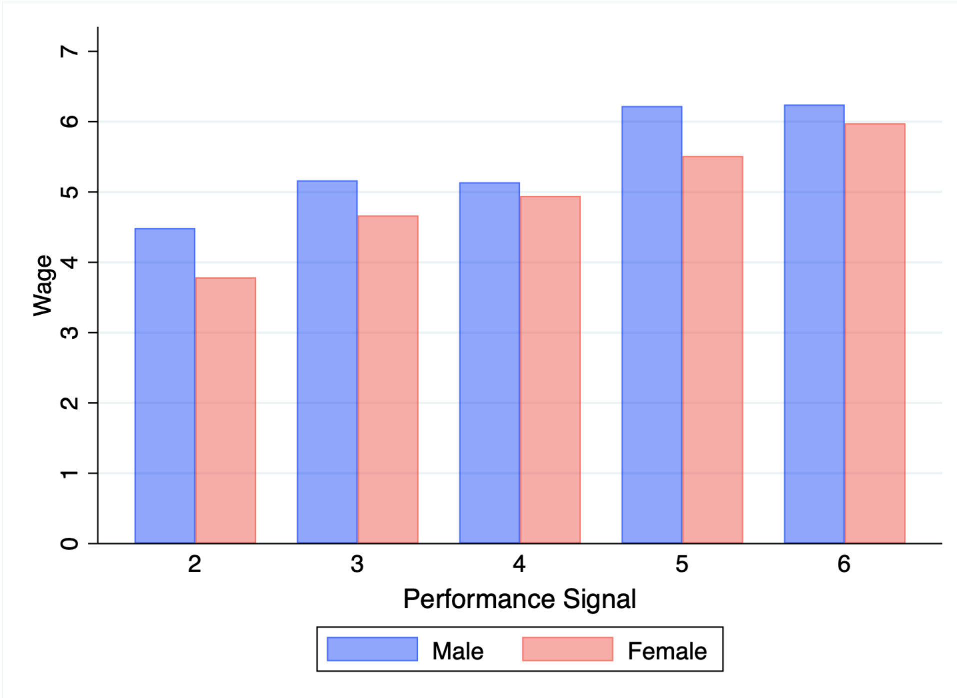

Figure 1. Signal Inflation: Recruiter Wage Offers by Worker Gender and Signal

Figure 1 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.

The direct discrimination in Figure 1 can be viewed as a measure of institutional discrimination among the "firm" of experimental Recruiters. To study whether such institutional discrimination is driven by in-group favoritism, we estimate the gender effect as a function of the Recruiter's own gender. Specifically, we estimate a regression of Recruiter wage offers on the Worker's gender separately for male and female Recruiters. Both sets of Recruiters offer significantly lower wages to female versus male workers. The coefficient for female Recruiters (0.30) is roughly half that of male Recruiters (0.68), though the difference between these coefficients is not statistically significant ($p = 0.12$). Changing the composition of female and male Recruiters may thus decrease overall institutional discrimination, though the effects of such compositional changes may not be significant.

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). Note that while Recruiter

discrimination in similar settings (Bordalo et al. 2019; Bohren et al. 2019).

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. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$;

signals (Task A performance) are informative for their objective (Task B performance), Hiring Manager signals (prior wage offers) may be judged to be more informative since they are based on the *same* performance information that Hiring Managers will be compensated on (Task A performance). Indeed, regressing Hiring Manager wage offers on Hiring Manager signals yields a larger coefficient (0.71) than with Recruiters.[44]

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 ($p < 0.01$). 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 thus 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 5.3. 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 A performance

[44]Since each Hiring Manager made only one offer, standard errors are not clustered in these analyses.

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 in Equation (3): $\delta(f, y)$. 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 discrimination against female Workers ranges from 0.47 to 2.01 for Task A 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 A 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 A 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 Systemic Discrimination and Screening

6.2.1 Design

Our second experiment examined the role of screening in generating systemic discrimination. The study also recruited three types of participants—Workers, Recruiters, and Hiring Managers—through the Prolific platform.

Workers: We used the same 100 Workers from the first experiment. Here we restricted attention to Workers with Task A performance in $\{2, 3, 4, 5\}$ in order to obtain enough data for each gender to be evaluated by Recruiters.

Recruiters: As in the first study, around 200 participants were randomized into the role of Recruiter. Recruiters were shown the Task A performance of two Workers and the Workers' gender. Thus, as before, Task A performance is the Recruiter's signal S_i^R, with $\mathcal{S}^R = \{2, 3, 4, 5\}$.

The main difference between Recruiters in the first and second study is the type of decision they were asked to make. Rather than stating their highest willingness to pay, Recruiters here were asked to select which of the two Workers they would like to hire. Recruiters 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 ($N = 506$).

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: As in the first study, we recruited around 500 participants for the role of Hiring Manager. Hiring Managers saw one Worker's profile who had been evaluated by a Recruiter, along with information about their gender. There were two key differences relative to the first study: in how the Hiring Managers' information was generated and decision environment. First, Hiring Managers 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. Second, rather than stating a wage, 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 in $S^H = \{2, 3, 4, 5\}$, 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. The relevant measure of Worker productivity for Hiring Managers is Task B performance, $Y_i^{H*} \in \{0, \ldots, 10\}$.

6.2.2 Results

As before, we measure systemic and total discrimination with respect to Task A performance, $Y_i^{R0} = Y_i^{H0} = S_i^R$, with $\mathcal{Y}^0 = \{2, 3, 4, 5\}$ here. 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%.[45] Given the lack of gender-based performance differences, as reported in Section 6.1.2, this disparity in hiring rates is not consistent with accurate statistical discrimination. Therefore, Recruiter direct discrimination again stems from either biased preferences or beliefs.

As in the first study, we can examine the extent of institutional discrimination in the "firm" consisting of our experimental Recruiters. We again find that both male and female Recruiters are less likely to hire female versus male Workers. Here, the coefficient for female Recruiters (0.25) is approximately 20% smaller than that of male Recruiters (0.31). The difference in coefficients in not statistically significant ($p = 0.66$), but does further illustrate how gender composition can impact the extent of institutional discrimination.

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 A 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 2 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

[45]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.

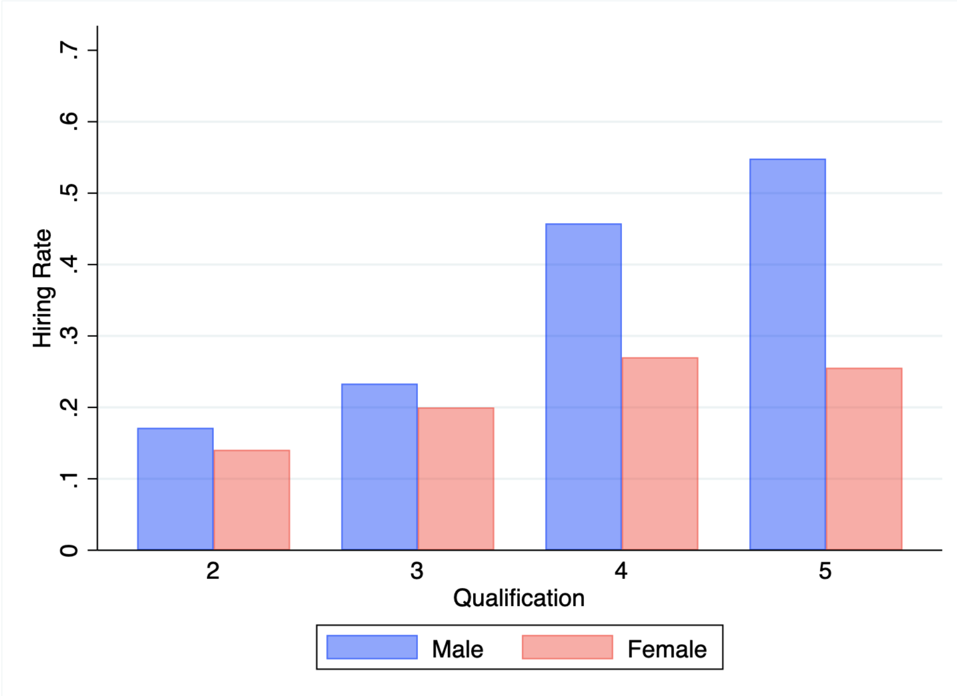

Figure 2. Screening: Manager Hiring Rates by Worker Gender and Qualification

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. Thus, systemic discrimination from screening hurts high-performing women the most.

As before we estimate systemic discrimination in Hiring Manager actions using the decompositions in Section 5.3. We first estimate total Hiring Manager discrimination $\varDelta(y^0)$ by comparing male and female hiring rates for each level of Task A performance $y^0 \in \mathcal{Y}^{H0} = \{2, 3, 4, 5\}$. 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 A 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 in equation Equation (3): $\delta(f, y)$. 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. Screening: Total, Direct, and Systemic Discrimination in Hiring Manager Actions

	(1)	(2)	(3)	(4)
	Task A 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 A. 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 ($N = 501$).

Table 4 confirms the heterogeneity in systemic discrimination faced by women with different productivity levels. At the two lower levels of Task A 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 2 comes from both types of discrimination in Hiring Manager actions.

In summary, our 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 correspondence and audit studies would not have detected the majority of discrimination in Hiring Manager wage offers or hiring rates: direct Hiring Manager discrimination, which conditions on the non-gender signal, was 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). Thus 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.

7 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 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.[46] 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.

[46]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 (2021a): "Measuring Racial Discrimination in Algorithms," *AEA Papers and Proceedings*, 111, 49–54.

——— (2021b): "Measuring Racial Discrimination in Bail Decisions," *NBER Working Paper No. 26999*.

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.

DellaVigna, S. and D. Pope (2018): "What Motivates Effort? Evidence and Expert Forecasts," *The Review of Economic Studies*, 85, 1029–1069.

Enke, B. and T. Graeber (2019): "Cognitive Uncertainty," *NBER Working Paper No. 26518*.

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.

Frydman, C. and L. J. Jin (2022): "Efficient Coding and Risky Choice," *The Quarterly Journal of Economics*, 137, 161–213.

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.

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.

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., 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.[47] 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 $(\mu_m - \mu_f)/(t+1)$.[48] Intuitively, the initial accurate statistical discrimination

[47] 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.

[48] 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, $F_s^R(s^R|y,m) = F_s^R(s^R|y,f)$. 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).