NBER WORKING PAPER SERIES

SYSTEMIC DISCRIMINATION:
THEORY AND MEASUREMENT

J. Aislinn Bohren
Peter Hull
Alex Imas

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

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

We thank Jason Abaluck, Amanda Agan, David Arnold, Marianne Bertrand, Peter Blair, Leonardo Bursztyn, Hanming Fang, Damon Jones, Diag Davenport, Stefano DellaVigna, Will Dobbie, Ed Glaeser, Sam Hirshman, Larry Katz, Erzo Luttmer, Matthew Knepper, Matt Lowe, Muriel Niederle, Matthew Notowidigdo, Anna Gifty Opoku-Agyeman, Jane Risen, Heather Sarsons, Andrei Shleifer, Fabio Tufano, Damián Vergara, Basit Zafar, and seminar participants and conference participants at various institutions for helpful comments. Joshua Hascher, Christy Kang, Yier Ling, and Cuimin Ba provided expert research assistance. Hull gratefully acknowledges funding through the National Science Foundation (Award #2119849). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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

ABSTRACT

Economics often defines and measures discrimination as disparities arising from the direct effect of group identity. We develop new tools to model and measure systemic discrimination, which captures how discriminatory decisions in other domains—past, future, or contemporaneous—contribute to disparities in a given decision. We show that systemic discrimination can be driven by disparate signaling technologies or differential opportunities for skill development. We then propose a new measure based on a decomposition of total discrimination into direct and systemic components, and show how it can be used to estimate systemic discrimination in both experimental and observational data. We illustrate these new tools in three applications, including a novel Iterated Audit experimental paradigm with real hiring managers. The applications also identify behavioral frictions that blunt the impact of individual-level interventions and perpetuate systemic discrimination, suggesting the need for systems-based policy responses to systemic discrimination.

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

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

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

1 Introduction

Disparities by race, gender, and other protected characteristics have been widely documented in many important settings—such as employment, housing, criminal justice, education, and healthcare.[1] In economics, research on such disparities tends to focus on the possibility of *direct discrimination*: differential treatment based on the protected characteristic itself, holding fixed all other observable characteristics. For example, a study of employer discrimination might examine racial disparities in job application callback rates while holding fixed all other attributes on the application (e.g., education and employment history).

A large body of work across other fields, however, suggests this view is incomplete. Sociologists and legal scholars have long emphasized the importance of a systems-based approach to study disparities, in which discrimination emerges as the cumulative outcome of both direct and indirect interactions across different time periods and domains (Pincus 1996; Powell 2007; De Plevitz 2007; Small and Pager 2020). More recently, computer scientists have shown large disparities in algorithmic treatment can arise indirectly from biased data collection and training systems, even when the algorithms are "group-blind"—i.e. not directly using protected characteristics (Angwin, Larson, Mattu, and Kirchner 2016; Rambachan and Roth 2020). From these perspectives, analyses of direct discrimination that condition on all observable characteristics fail to capture the full scope of inequity: such characteristics may themselves "bake-in" discrimination from interactions with other individuals, institutions, and markets.[2] As Small and Pager (2020) write: "Though [canonical economic models] make different assumptions about why discrimination happens, they all agree on a core issue: for discrimination to happen, an individual must decide to treat people of different backgrounds differently [...] As a result, they miss what sociologists and others have called 'institutional discrimination,' 'structural discrimination,' and 'institutional racism,' which are all terms used to refer to the idea that something other than individuals may discriminate."[3]

This paper develops a general framework nesting these different forms of discrimination in a common structure, and introduces new tools to bring this framework to data. We make three main contributions: a theoretical framework for formally defining and exploring

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

[2]Some versions of this "bad control" problem—i.e., conditioning on characteristics that are themselves affected by discrimination—have also been studied in economics, most prominently as "premarket factors" in studies of labor market disparities (e.g., Neal and Johnson (1996), Altonji and Blank (1999), and Coate and Loury (1993)). The framework in this paper connects these earlier studies to broader notions of systemic discrimination and allows for general analyses of sources and policy responses, as well as measurement.

[3]They also note that "unfortunately, these terms are not used consistently across the social sciences; moreover, they are often used even more ambiguously among lay writers and commentators." Our paper seeks to address this by providing a unified and precise framework to study these broader notions of discrimination.

broader notions of discrimination that account for systemic factors, an empirical framework for isolating the impact of systemic factors by decomposing discrimination into direct and systemic components, and three empirical applications. The applications demonstrate our new empirical tools, show the large impact that systemic factors can have on disparities both over time and across markets, and identify key psychological frictions leading to persistent systemic discrimination. Overall, this analysis yields new insights for the interpretation and measurement of widely-documented disparities and potential policy responses.

To illustrate the importance of a systems-based approach to studying discrimination, consider the case of *Griggs v. Duke Power Co. (1970)*: a landmark Supreme Court decision on the interpretation of Title VII of the U.S. Civil Rights Act. Griggs argued that Duke Power's policy of requiring a high school diploma for within-company transfers was discriminatory because it disadvantaged Black employees who were qualified but lacked a degree, in part due to ongoing discrimination in secondary education. The Court agreed, noting that the degree requirement bore no relevance to an individual's ability to perform different jobs at the company. Notably, discrimination was found despite the policy being facially "race-blind": white and Black employees with the same educational background had the same ability to transfer jobs. Standard measures of discrimination that condition on educational background would thus have failed to capture the discrimination found among Black and white workers with the same qualification (i.e., the ability to perform a specific job). Canonical models of taste-based and statistical discrimination (Becker 1957; Phelps 1972) are similarly inappropriate for understanding the source of this indirect form of discrimination.

To develop a systems-based approach, we first present a theoretical framework for comparing and contrasting the canonical notion of direct discrimination in economics to broader notions of discrimination from other fields. We formalize a notion of *total* discrimination—group-based disparities among equally-qualified individuals—which captures both direct and systemic factors. *Direct* discrimination reflects the differential treatment of protected groups on the basis of group membership itself, while *systemic* discrimination captures disparities among equally qualified individuals that arise indirectly when interactions in other periods or contemporaneously across domains generate differences in relevant non-group characteristics (e.g., a recommendation letter or test score). To illustrate these definitions in the case of *Griggs*, there was no direct discrimination since the company's policy was facially race-blind. Instead, there was systemic discrimination stemming from racial differences in the distribution of high school diplomas among equally productive workers.

We use this theoretical framework to delineate two sources of systemic discrimination, akin to the canonical typology of direct discrimination as being taste-based or statistical. *Informational* systemic discrimination arises from group-based differences in the distribution of signals of the payoff-relevant attribute (e.g., test scores as a signal of productivity) among similarly qualified individuals. We show how such differences can stem from signal inflation, in which signals are systematically higher for one group (e.g. bias due to past discrimination), or in screening decisions when signals are more informative for one group than another (e.g., data availability differs across groups due to systemic barriers). *Technological* systemic discrimination instead arises from differences in the distribution of the payoff-relevant attribute itself. For example, two workers with the same initial human capital may have systematically different subsequent productivities due to discrimination in their access to intermediate employment and schooling.

One key practical takeaway of this framework is that well-posed studies of discrimination require researchers to take an explicit stance on the notion of "equally-qualified," which amounts to a choice of reference qualification they measure discrimination with respect to. Focusing on direct discrimination implicitly takes the stance that individuals from different groups are equally-qualified if they match on other observable characteristics—such as educational attainment, in the case of *Griggs*. Other choices of reference qualification, such as worker productivity or initial human capital, allow for the study of different systemic forces. Through different choices of reference qualification, the framework provides a unified structure nesting different forms of discrimination. In any given setting, there may be one or several natural choices depending on which systemic forces are of interest to the researcher. Our framework and empirical applications highlight the importance of making this choice explicit for interpreting observed disparities and forming appropriate policy responses.

We next develop an empirical framework for bringing these concepts to data. We derive measures of systemic discrimination from a novel Kitagawa-Oaxaca-Blinder (KOB) decomposition of total discrimination into direct and systemic components.[4] Direct discrimination can be identified through conventional audit or correspondence studies, which measure the causal effect of perceived group membership on an action holding fixed all other observable characteristics. Total discrimination is identified by disparities that condition on the researcher-chosen qualification measure. Our KOB decomposition shows how these measures can be combined to identify systemic discrimination.

An important takeaway from the empirical framework is that different types of (quasi-) experimental data and designs may be needed to quantify systemic discrimination, relative to established methods for measuring direct discrimination. We present a general experimental approach, termed an *iterated audit* (IA), which can be used to measure systemic discrimination when the researcher-chosen qualification is observed. We also discuss strategies for identifying or bounding systemic discrimination when the qualification is only selectively

[4]KOB decompositions (named after Kitagawa (1955); Oaxaca (1973); Blinder (1973)) are typically used to measure direct discrimination as the residual of an unconditional disparity after accounting for differences in observable characteristics. Our decompositions instead measure systemic discrimination as the residual of a measure of total discrimination after accounting for a measure of direct discrimination.

observed, or when it is unobserved but can be reliably predicted by observables.[5]

We illustrate these new tools in three empirical applications. First, we use the iterated audit approach in a lab-in-the-field experiment to measure systemic discrimination in hiring propensities. We recruited real hiring managers and asked them to evaluate a series of resumes for entry-level jobs using an incentivized ratings design (Kessler, Low, and Sullivan 2019). Based on results from a previous audit study by Pager (2003)—who found Black applicants were significantly less likely to proceed through an entry-level job application process than equally-qualified white applicants—we generated two distributions of entry-level work experience: one commensurate with the rate at which white workers proceeded through the application process and one commensurate with the rate for Black workers. We then generated three sets of resumes. Two sets were as in a standard correspondence study, with set A assigned distinctively white names, set B assigned distinctively Black names, and both assigned entry-level work experience drawn from the white distribution. The third set (C) was assigned distinctively Black names and entry-level work experience drawn from the Black distribution. Aside from entry-level work experience and white versus Black names, all three sets had similar resume characteristics. Set C incorporated earlier direct discrimination, since it has a different distribution of entry-level work experience as documented in Pager (2003). Hiring managers evaluated resumes on the likelihood of hiring the applicant on a scale of 1 to 10. Importantly, managers were also informed of the previous Pager (2003) study's findings and how the disparities in work experience were generated.

The IA design reveals significant racial discrimination in hiring propensities, driven primarily by systemic discrimination from differential work experience. Comparing evaluations of resumes in sets A and C identifies total discrimination: resumes in set C had a substantially lower hiring likelihood—roughly 20% of a standard deviation—compared to resumes in set A. Comparing sets A and B identifies the direct discrimination component, which was small and insignificant. The residual comparison of sets B and C identifies the systemic discrimination component, as it compares the resumes that Black workers would have had in the absence of prior discrimination to resumes that Black workers actually have given prior discrimination—hence, it isolates the impact of past discrimination on the current evaluation. Systemic discrimination was large and comprised the vast majority of total discrimination. Importantly, systemic discrimination arose despite hiring managers being made

[5] Several recent papers have applied our framework to quantify systemic discrimination. Baron, Doyle Jr, Emanuel, Hull, and Ryan (2023) use quasi-experimental variation to study how racial discrimination propagates through different phases of the Michigan child welfare system. Zivin and Singer (2023) study racial differences in home values as a function of pollution exposure, documenting a large disparity of which 75% is attributable to systemic discrimination through complementary amenities. Gawai and Foltz (2023) look at the impact of country of birth on income in academia and show that 67% of the disparity is driven by systemic discrimination. See also Althoff and Reichardt (2022), Lodermeier (2023), Buchmann, Meyer, and Sullivan (2023), and Conway, Mill, and Stein (2023).

aware of past direct discrimination before forming their evaluations.

These results suggest behavioral frictions may limit the effectiveness of information on past direct discrimination for mitigating the effects of systemic discrimination. Outcome bias—a phenomenon in which evaluators over-attribute variation in relevant outcomes (e.g., work experiences) to differences in personality factors (e.g., effort or ability) while under-attributing contextual factors (e.g., luck or previous evaluation bias)—has been identified as an important friction blunting the effectiveness of information in other domains (e.g. Brownback and Kuhn (2019)). Our second lab experiment directly explores how such frictions interact with systemic discrimination. One set of participants, Workers, were randomized into one of two groups and completed a series of tasks. A second set of participants, Recommenders, were given Worker profiles and incentivized to correctly predict the number of completed tasks based on a noisy signal of performance. The signal was either biased upward or downward based on the Workers' randomly-assigned group membership. A third set of participants, Evaluators, were also incentivized to predict Worker performance but observed the Recommender assessment rather than the biased signal. Importantly, both Recommenders and Evaluators were aware of the signal bias and the fact that Worker group was randomly assigned (and thus unrelated to Worker productivity). We studied Recommender and Evaluator assessments in two regimes: a Group-Blind treatment in which they were unable to account for bias in the signal and past assessments, and an Informed treatment in which they were able and explicitly incentivized to offset the bias.

As in the first experiment, we find that information on past direct discrimination does little to blunt the impact of systemic discrimination—despite the clearly structured incentives and transparency in the signal-generating process. We observe large group-based disparities in the assessment of both Recommenders and Evaluators in the Group-Blind treatment. Taking true productivity as the reference qualification, these disparities can be interpreted as total discrimination. In the Group-Blind treatment there is no scope for direct discrimination, meaning all of this discrimination is systemic (arising from signal inflation). Interestingly, however, we find similarly large levels of total discrimination in the Informed treatment—despite Recommenders and Evaluators knowing exactly how the signals are biased and that group membership is arbitrary. Informed Recommenders offset some of the signal bias, but most discrimination remains; Informed Evaluators make no statistically significant adjustment for Worker group in their assessments. Recommenders thus appear to exhibit outcome bias in over-attributing group-based differences in the signal to productivity differences, rather than the known signal bias, while Evaluators appear to fully discount the possibility of any remaining bias in the Recommenders' assessments.

Our third experiment shows that systemic discrimination is not limited to dynamic processes and can arise through contemporaneous decisions across domains. Specifically, we document technological systemic discrimination driven by the interaction between two contemporaneous decisions. Participants were randomly assigned to one of two groups and chose how much effort to exert. They were paid by two companies based on a noisy signal of this effort: one company paid members of one group systematically less than the other, conditional on the same effort choice, while the other company did not directly discriminate. We find that participants who anticipate direct discrimination from the first company choose to put in less effort. This led to group-based disparities in the payments from the second company, despite it not engaging in any direct discrimination. Since there were no initial group-based differences in qualifications by design, this disparity was driven by systemic discrimination that emerged as a result of anticipated direct discrimination.

Taken together, our findings demonstrate the importance of a systems-based approach for studying and addressing discrimination. As in *Griggs*, the majority of inequity found in each of the three applications would be missed by a conventional analysis of direct discrimination. Moreover, the potential for outcome bias and anticipatory effects suggest simple interventions that focus on individual decision-making (e.g. providing evaluators with information) may have little effect on mitigating total discrimination. Instead, large and pervasive systemic discrimination may necessitate broader or system-based policy responses.

We organize the rest of this paper as follows. Section 2 presents a simple example motivating a systems-based approach to discrimination. Section 3 develops our theoretical framework for studying direct and systemic discrimination. Section 4 discusses measurement, including our decomposition of total discrimination into direct and systemic components and the IA design. Section 5 presents the applications. Section 6 concludes. In Appendix A we review connections to related literatures and in Appendix B we present additional applications.

2 Motivating Example

We begin with a simple theoretical example that illustrates the key features of the framework.[6] Consider a population of patients i seeing a physician in order to decide whether to get a colorectal cancer screening. The goal of such screenings is to detect early signs of cancer. Correspondingly, let $Y_i^* \in \{0, 1\}$ indicate the latent presence of cancer in patient i.

Before seeing the physician, each patient is first seen by a nurse practitioner (NP). The NP takes down the patients demographics (including self-reported race, G_i), checks their basic medical information (e.g. height, weight, blood pressure), and conducts a short medical history survey. The survey includes a variety of open-ended questions on the patient's experience with screening and cancer, and NPs have some discretion as to how they record a patient's answers. The physician receives a file from the NP with all the collected information, conducts their own short interview, then makes a screening recommendation. Let S_i denote

[6] This example is inspired by Zink, Obermeyer, and Pierson (2023).

all information available to the physician for patient i excluding the self-reported race, and let A_i denote her action (a screening recommendation).

A large literature has found substantial race-based disparities in screening decisions (e.g., Jerant, Fenton, and Franks (2008), Crawley, Ahn, and Winkleby (2008)); three economists are interested in studying the role of discrimination in explaining these disparities. Economist 1 follows standard practices in the field by designing and conducting a careful audit study. Specifically, she recruits a set of white and Black patients with comparable demographics and basic medical information, and randomizes them to pre-filled NP files and scripts for interacting with the physician. In this way, she ensures identification of the effect of race G_i on the action A_i conditional on the physician's non-race signals S_i. She finds that white patients are, on average, somewhat less likely to receive a screening recommendation than Black patients assigned to the same file and script. She concludes that there is some racial discrimination against white patients in this setting.

Economist 2 is interested in the same question, but ends up running a somewhat different audit study. Rather than randomizing NP files directly, she randomizes scripts for the recruited white and Black patients to interact with *both* the NP and physician. That is, while she ensures white and Black patients have the same screening and cancer history to report to the NP, she allows NPs to affect the recording of this information that is given to the physician through S_i. Strikingly, this disparity in the design yields a different conclusion than Economist 1's: white patients are, on average, slightly *more* likely to receive a screening recommendation than Black patients assigned to the same set of scripts. Thus she concludes there is some racial discrimination against Black patients in this setting. In unpacking this result, she finds that NPs tend to use more serious language in recording the history of white patients relative to Black patients randomized to the same family history.

Finally, Economist 3 examines the same question by running a different type of study. Randomly screening a representative set of white and Black patients after the physician makes a recommendation, she measures true rates of cancer incidence Y_i^*. This allows her to compute racial disparities in screening recommendations among patients with the same cancer status, without conditioning on any non-race signals. Curiously, she reaches a different conclusion than both Economist 1 and 2: white patients are *much* more likely to receive a screening recommendation than Black patients with the same underlying cancer status. In unpacking this result, Economist 3 finds that a key driver is the differential accuracy of available family history information by race: among patients with the same cancer status, Black patients are much less likely to know whether their parents or grandparents suffered from colorectal or related cancers due to more limited historical interactions with doctors.[7]

At first blush, this simple example presents a puzzle: which of the three researchers are

[7] See, e.g., Kupfer, McCaffrey, and Kim (2006).

correct on the nature and extent of discrimination in cancer screening recommendations? Economist 1 follows the norm in economics by conditioning on the information available to the decision-maker, S_i. Economist 2 finds that some of this signal is biased by a different stage of the recommendation system (i.e. the language disparity in the NP's notes); her measure of discrimination takes a starting point before the patient's current interaction with the healthcare system. By conditioning on an "objective" measure of qualification for screening—the underlying cancer risk Y_i^*—Economist 3 uncovers a further bias in the patient's recorded family history. How can such *systemic* biases be coherently studied alongside the *direct* discrimination that Economist 1 documents in physician decision-making?

In Section 3 we develop a general framework for reconciling these different analyses. The framework formalizes how the study of discrimination requires a researcher to take a stance on the notion of "qualification" for a given decision—what we call Y_i^0—which is the key factor that differed across the three researchers. In any given setting, there may be one or several natural choices for Y_i^0; by selecting different reference qualifications, a researcher can study different systemic forces alongside canonical sources of direct discrimination.

Two other points, which we return to in subsequent sections, are worth highlighting in this example. First, as shown in Section 4, studying such systemic forms of discrimination generally requires new empirical tools. While Economist 1 identified direct discrimination with a standard audit experiment, isolating the effects of the systemic biases found by Economists 2 and 3 is more challenging. We propose an alternative iterated audit design to identify these effects, and discuss how applying this design may require different (quasi-) experimental designs—particularly when the chosen qualification reference point is imperfectly observed.

Second, as we illustrate with empirical applications in Section 5, it may be difficult to address such systemic forms of discrimination with standard individual-level interventions. The physician in this example seems at least partly aware of the systemic bias in NP notes, given her "reverse" discrimination in recommending screenings at a lower rate for white patients than observably-similar Black patients. Yet she does not fully offset the bias, either because of imperfect awareness or because of her own psychological frictions or biases. Our applications show such frictions can be very important, even in settings where systemic discrimination is transparent and decision-makers have an explicit incentive to counteract it. Hence broader or system-wide policy responses may be called for in settings with significant systemic discrimination. By nesting different forms of discrimination in a single framework, our approach can be used to formulate and target such systems-wide policy responses and to study how they may impact other interconnected decisions.

3 Formalizing Systemic Discrimination

We now develop our general theoretical framework to compare and contrast the canonical notion of direct discrimination in economics to broader notions of discrimination from other fields. Section 3.1 introduces the setting and Section 3.2 defines systemic and total discrimination. Section 3.3 discusses key features of the framework and Section 3.4 enriches it by delineating two main sources of systemic discrimination.

3.1 Preliminaries

We develop our framework in a labor market context, in which a manager evaluates workers for a task. Worker i has ex-ante unobservable *productivity* $Y_i^* \in \mathcal{Y}^*$, where $\mathcal{Y}^* \subset \mathbb{R}$ is the set of possible productivity levels.[8] The manager first observes the worker's *group* identity $G_i \in \{b, w\}$ and a vector of k attributes $S_i \in \mathcal{S}$ (e.g., educational background, prior evaluations, etc.) which we refer to as the *signal*, where $\mathcal{S} \subset \mathbb{R}^k$ is the set of possible signal realizations. A worker's attributes and group identity can potentially provide information about the worker's productivity. The manager then takes an *action* $A_i \in \mathcal{A}$, where $\mathcal{A} \subset \mathbb{R}$ is a finite action space.[9] This could be a binary decision such as hire or don't hire, a wage, or another type of evaluation (e.g., a multi-valued rating). The manager's payoff depends on productivity, group identity, and her action; specifically, she maximizes expected utility subject to her beliefs about the joint distribution of productivity, the signal, and group identity. Rather than explicitly modeling the manager's decision problem, we take a reduced-form approach by specifying the manager's decision rule $A : \{b, w\} \times \mathcal{S} \to \mathcal{A}$, which determines how her information set (i.e., observed signal and group identity) maps into an action choice. Given G_i and S_i, the manager selects action $A_i = A(G_i, S_i)$ for worker i.[10]

We embed the employment evaluation in a broader economy to capture the idea that a worker's productivity and signal in the current evaluation may be affected by decisions in other markets and time periods: i.e., *systemic* forces. Worker i enters the economy with *qualification* $Y_i^0 \in \mathcal{Y}^0$, where $\mathcal{Y}^0 \subset \mathbb{R}$ is the set of possible qualification levels. The qualification serves as a reference point from which one can measure how the accumulation of subsequent systemic forces impact the current evaluation. Importantly, as the motivating example shows, it is a choice variable of the researcher. Possible choices include productivity in the current task, $Y_i^0 = Y_i^*$, the signal in the current task, $Y_i^0 = S_i$, or an earlier reference point such as productivity or available information earlier in the pipeline. In this third case, current productivity and the signal may arise endogenously from the chosen qualification and

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

[9] We assume the action space is finite to simplify notation. The extension to a continuous space (e.g. $\mathcal{A} = \mathbb{R}$) is immediate, with the addition of the appropriate measure-theoretic statements to define probabilities when any individual element of a set has zero probability.

[10] We abstract from interactions across workers and other realistic features of labor markets for simplicity.

actions subsequent to this reference point. Setting Y_i^0 equal to a constant selects a reference point prior to the emergence of any group differences. Regardless of the chosen qualification, we emphasize that it need not represent a fixed or "inherent" characteristic of the worker; it is a reference point that fixes initial conditions in a specific context. We return to this point in Section 3.3, showing how different choices of qualification allows us to nest different notions of discrimination within a unified framework.

It is straightforward to map this framework to the example in Section 2 studying discrimination in cancer screening recommendations A_i. Economist 1 selects the physician's signal as the qualification, $Y_i^0 = S_i$; this includes the nurse practitioner's action. Economist 2 selects the nurse practitioner's signal as the qualification. Economist 3 selects the patient's cancer status as the qualification, $Y_i^0 = Y_i^*$. It is also straightforward to map this framework to other settings: for example, in a lending context, a loan officer may decide whether to lend to borrowers (A_i) who differ in their ability to pay back the loan (Y_i^*) and who report signals (S_i) of creditworthiness such as credit scores and income. A researcher may select as the qualification these signals, true creditworthiness, or some initial lending qualification.

In these and other examples, the chosen qualification Y_i^0 can interact with decisions made in other markets and domains to determine how Y_i^* and S_i vary by group identity G_i. Some of these differential interactions may arise from the direct discrimination typically considered in economics. The accumulation of such interactions leads to a broader notion of discrimination, which we formalize next.

3.2 Definitions

We define three forms of discrimination in the manager's action with respect to worker group: direct, total and systemic. *Direct discrimination* captures group-based differences in manager actions, holding fixed the productivity signal. It occurs when the action rule prescribes different actions for group w and b workers with the same signal realization:

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

Direct discrimination arises from the worker's group identity itself; it is a causal concept because it conditions on all observed non-group attributes. It can arise from the dependence of the manager's preferences, beliefs about productivity, or beliefs about the signal distribution on group identity.

Our definition of *total discrimination* captures a broader notion of inequity that incorporates how decisions in other domains and markets contribute to disparities in the present one. Let $\mu^g(a; y^0)$ denote the probability of action a for workers of group g and qualification level y^0. Total discrimination occurs when group w and b workers with the same qualification level face different action distributions:

Definition 2 (Total Discrimination)**.** *The manager exhibits total discrimination at qualification level $y^0 \in \mathcal{Y}^0$ if $\mu^b(a; y^0) \neq \mu^w(a; y^0)$ for some $a \in \mathcal{A}$.*

Discrimination in decisions that occur subsequent to the chosen qualification will be accounted for in total discrimination. This includes decisions prior to the current decision that directly impact the current signal and productivity, as well as contemporaneous and future decisions where the prospect of discrimination in these decisions influences the current signal and productivity. For example, anticipating a discriminatory jury (the future decision) impacts the evidence a defense attorney acquires (the current signal) to present to the judge for a plea deal (the current evaluation).

The possibility of discrimination in separate decisions driving discrimination in the current one motivates our final definition. Following Pincus (1996) and Gynter (2003), *systemic discrimination* corresponds to disparities in manager actions that stem from discrimination in other decisions, i.e., from parts of the system separate from the current evaluation. While total discrimination incorporates such disparities, it also includes disparities arising from direct discrimination in the present task.[11] To isolate this systemic component, we shut down direct discrimination by considering how the action distribution varies by group in the counterfactual situation where both groups face the same action rule for this current evaluation. Formally, let $\sigma^g(s; y^0)$ denote the probability of group g workers with qualification level y^0 realizing a set of signals s. Note that the action distribution at y^0 can be expressed as the probability of the set of signals that map to action a under the action rule for group g:

$$\mu^g(a; y^0) = \sigma^g(\{s : A(g, s) = a\}; y^0). \tag{1}$$

First, suppose group b faced the group w action rule, $A(w, s)$. Define a counterfactual action distribution for group b as the action distribution under this action rule and the signal distribution for group b,

$$\tilde{\mu}^b(a; y^0) \equiv \sigma^b(\{s : A(w, s) = a\}; y^0). \tag{2}$$

In words, this counterfactual action distribution captures how group b workers would be evaluated if they were subject to the same treatment as group w workers with the same signal realization and reference qualification. Analogously, define a counterfactual action distribution for group w under the action rule for group b as $\tilde{\mu}^w(a; y^0) \equiv \sigma^w(\{s : A(b, s) = a\}; y^0)$. Comparing the counterfactual action distribution for group b to the actual action distribution for group w, or vice versa, determines whether decisions in other domains impacted the signal and productivity in a way that led to different treatment of group w and b workers with the same reference qualification—that is, systemic discrimination:

[11] This is because the action distribution depends on the action rule, as can be seen in Eq. (1) below.

Definition 3 (Systemic Discrimination)**.** *The manager exhibits systemic discrimination at qualification level* y^0 *if* $\mu^w(a; y^0) \neq \tilde{\mu}^b(a; y^0)$ *or* $\mu^b(a; y^0) \neq \tilde{\mu}^w(a; y^0)$ *for some* $a \in \mathcal{A}$*.*

As in the case of total discrimination, our definition of systemic discrimination conditions on qualification and hence captures disparities stemming from systemic forces that occur subsequent to that chosen reference point. As we discuss more in Section 3.4, systemic discrimination can stem from group-based differences in the signal distribution for workers with the same productivity, group-based differences in the productivity distribution for workers with the same qualification, or both. Hence, this definition captures disparities arising from the impact of decisions in other domains on both the signal and productivity that are relevant for the current evaluation.

To illustrate these definitions we return to the example in Section 2. Consider a Black and white patient with the same screening and cancer history (the reference qualification of Economist 2), and suppose the nurse practitioner (NP) assigns each patient a risk score of low, medium or high (the doctor's signal). The nurse practitioner's bias leads the Black patient to receive systematically lower risk scores than the white patient. Suppose that doctors underscreen Black patients—specifically, they screen white patients when they receive a risk score of medium or high while Black patients are screened when they have a risk score of high. Direct discrimination captures disparities stemming from the different screening thresholds (medium versus high): Black patients have a higher threshold than white patients, and therefore Black and white patients with the same risk score are screened at different rates. Systemic discrimination captures disparities stemming from the bias in assigning a risk score. Fixing the screening threshold as that used for white patients, Black patients receive a risk score of medium or high with lower probability than equally-qualified white patients. Total discrimination aggregates both of these components: it compares the probability that a white patient receives a risk score of medium or high to the probability that an equally-qualified Black patient receives a risk score of high. This captures the total difference in the screening probability for equally-qualified white and Black patients.

3.3 Discussion

The Choice of Qualification Y_i^0. This framework makes clear that the study of discrimination requires a researcher to take a stance on the choice of qualification Y_i^0. The interpretation of systemic and total discrimination is inherently tied to this reference point since only disparities that emerge subsequent to the chosen qualification are included in these measures of discrimination. While prior work often makes this choice implicitly, more explicit discussion is critical for interpreting results and forming an appropriate policy response. In any given setting, there may be one or several natural choices for Y_i^0 depending on which forms of discrimination are of interest to the researcher.

At one extreme, when qualification is set equal to the signal, $Y_i^0 = S_i$, total discrimination is narrowly defined as any group-based disparities that remain when holding fixed the relevant observables. Hence total and direct discrimination coincide and there is no scope for systemic discrimination. This is the choice of Economist 1 in Section 2, and is the implicit choice in most economic analyses of direct discrimination. At the other extreme, when worker qualification is set equal to a constant, $Y_i^0 = 0$, total discrimination corresponds to the unconditional disparity between groups. This choice thus yields the broadest measure of systemic discrimination, which accounts for any indirect relationship between group identity and the current productivity or signal.[12]

By selecting a reference point in between these two extremes, the researcher can study different systemic forces in the economy. A focal case sets the qualification to productivity, i.e. $Y_i^0 = Y_i^*$, the payoff-relevant outcome, e.g., Economist 3 setting Y_i^0 to the underlying risk of cancer. Total discrimination then corresponds to treatment differences for workers with the same productivity. For example, suppose a training program or club membership acts solely as a signaling device and has no impact on the manager's payoff. Setting productivity as the reference point, total discrimination accounts for indirect discrimination stemming from differential access to the signaling opportunity, whereas direct discrimination does not.[13] Notably, choosing productivity as the qualification aligns total discrimination with the legal notion of *disparate impact*, as it allows for disparities relevant to "business necessity." For example, Arnold et al. (2022), consider a measure of disparate impact in the pretrial setting where $Y_i^0 = Y_i^*$ is pretrial misconduct potential. This case also aligns total discrimination with some measures of algorithmic unfairness, where the action is a prediction of some unobserved state Y_i^* (e.g., Berk, Heidari, Jabbari, Kearns, and Roth 2018).

Another alternative selects productivity at a prior decision as the qualification. This yields a measure of discrimination that accounts for how other decisions—including the worker's own—impact current productivity Y_i^*. For example, a worker's employment history may impact her current labor market productivity Y_i^*. To study the impact of employment history, the researcher could choose productivity when entering the labor market as the qualification; total discrimination then corresponds to treatment differences in the present hiring task for workers with the same initial labor market productivity. Similarly, selecting available information at a prior decision as the qualification accounts for how other decisions impacted current information S_i. For example, Economist 2 in Section 2 selects the qualification as the information available to the NP rather than the physician, which accounts for

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

[13]This is the reasoning behind legal cases made against group-based exclusivity in country clubs, which offer members a host of pecuniary and non-pecuniary benefits such as access to networks (Jolly-Ryan 1998).

any systemic discrimination that stems from direct discrimination in the NP's decision (see Appendices B.1 and B.2.2 for an empirical illustration).

Importantly, the qualification metric can be chosen to include or exclude group-based differences in preferences that generate differences in productivity and the signal. For example, suppose racial or gender socialization affects the worker's decisions in a way that affects her work history or ability to signal her productivity (e.g., choosing a job with a flexible schedule, refraining from asking for a raise). Setting Y_i^0 upstream of such socialization captures this channel as systemic discrimination. Alternatively, including a measure of such preferences in Y_i^0 shuts down this channel (as in, e.g., Cook, Diamond, Hall, List, and Oyer 2021).

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. Importantly, as in the motivating example, it also allows for the interpretation of seemingly disparate findings and the study of policy interventions at different decision points.

Relation to Systemic Discrimination in Other Literatures. Our definition of systemic discrimination aligns broadly with how systemic and structural discrimination are discussed in the sociology literature: as a form of inequality operating indirectly through characteristics beyond group identity. 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)).[14] Correspondingly, in our definition, systemic discrimination can generate total discrimination even when there is no direct discrimination in the current evaluation—i.e., the action rule is group-neutral—because this group-neutral action rule fails to account for discrimination in other time periods or domains or even intentionally builds in discrimination indirectly by using other signals to proxy group.[15] Powell (2007) defines systemic discrimination as a "product of reciprocal and mutual interactions within and between institutions," both "within and across domains."[16] Similarly, our definition of systemic discrimination captures disparities that arise from the interaction between discriminatory decisions across time and contemporaneously across different domains. Systemic discrimination can emerge when past discriminatory decisions impact present decisions—so-called "past-in-

[14] For example, the historical practice of "redlining" in mortgage markets prioritized borrowers from majority-white neighborhoods over equally-creditworthy borrowers from majority-Black neighborhoods. Such neighborhood-based prioritization generated substantial race-based lending disparities despite the policy being prima facie race-neutral.

[15] Our definition also aligns broadly with what is sometimes referred to as institutional discrimination (Small and Pager 2020), though other types of institutional discrimination, e.g., when direct discrimination is codified into policy such as the case of Jim Crow laws, is a separate phenomenon.

[16] 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.

present" discrimination (Feagin and Feagin 1978), as illustrated in Section 5.1, Section 5.2, and Appendix B.1.[17] It can also can emerge across domains when discriminatory practices in one market impact productivity or signaling in another—so-called "side-effect" discrimination (Feagin and Feagin 1978), as illustrated in Section 5.3. We further review connections to the sociology literature on systemic discrimination, as well as notions of discrimination in law, economics, and computer science, in Appendix A.

3.4 Sources of Systemic Discrimination

Systemic discrimination arises from the interaction of group-based differences in the signal distribution at a given qualification, $\sigma^g(s; y^0)$, with how signals map into actions via the action rule. If the signal distribution is equal for both groups, $\sigma^b(s; y^0) = \sigma^w(s; y^0)$, then the actual and counterfactual action distributions are equal and there is no scope for systemic discrimination. If the signal distributions differ, then whether systemic discrimination arises depends on the interaction between the dependence of $A(g, s)$ on s and the way in which the signal distributions differ. Specifically, given a set of signals that map to the same action, the signal distributions must differ in terms of the probability assigned to this set.[18]

To delineate two main sources of systemic discrimination, we note that group-based differences in the signal distribution $\sigma^g(s; y^0)$ can be further split into two components: an *informational* channel stemming from differences in the signal of productivity for group b versus w workers with the same productivity, and a *technological* channel stemming from differences in productivity accumulation for group b versus w workers with the same qualification level. Formally, let $\sigma^g(s|y^*; y^0)$ denote the signal distribution for group g workers with productivity y^* and qualification level y^0, and let $\phi^g(y^*; y^0)$ denote the productivity distribution for group g workers with qualification level y^0. Note that $\sigma^g(s; y^0) = \int_{\mathcal{Y}^*} \sigma^g(s|y^*; y^0)\phi^g(y^*; y^0)dy^*$. Group-based differences in $\sigma^g(s|y^*; y^0)$ capture the informational channel, while group-based

[17] Past-in-present discrimination can also emerge when a system or institution is first "designed" by a group in power, which leads to the development of evaluation criteria that are optimized around the characteristics of this group. For example, De Plevitz (2007) discusses the impact of the "Eurocentric model of teaching" on schooling outcomes of Aboriginal children in Australia, noting 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. Another example is the practice of excluding women or minority groups from medical trials, which leads to a less informative signal of the efficacy of new treatments for these groups (Bierer, Meloney, Ahmed, and White 2022). In our framework, this corresponds to viewing the signal distribution as a choice variable for the dominant group, similar to the discussion in Pincus (1996).

[18] As an example, return to the illustration at the end of Section 3.2. Now suppose the doctor screens both groups following a medium or high risk score. If the risk score distributions for Black and white patients differ only in the relative probability of medium versus high but not the total probability of medium and high, then Black and white patients are screened with the same probability and there is no systemic discrimination. On the other hand, if Black patients are assigned a medium and high risk score with lower probability than white patients, then this difference *does* lead to systemic discrimination. Concretely, $(\sigma^b(m; y^0), \sigma^b(h; y^0)) = (0.3, 0.2)$ and $(\sigma^w(m; y^0), \sigma^w(h; y^0)) = (0.2, 0.3)$ differ but do not lead to systemic discrimination since $\sigma^b(\{m, h\}; y^0) = \sigma^w(\{m, h\}; y^0) = 0.5$. If instead $(\sigma^b(m; y^0), \sigma^b(h; y^0)) = (0.2, 0.2)$ then systemic discrimination arises since $\sigma^b(\{m, h\}; y^0) = 0.4$.

differences in $\phi^g(y^*; y^0)$ capture the technological channel. We discuss each 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. For example, borrowers with the same ability to repay (Y_i^*) and initial lending qualification (Y_i^0) may have credit histories (S_i) that are differentially informative due to discrimination in past borrowing opportunities (Bartik and Nelson 2016).

One salient form of informational systemic discrimination is *signal inflation*, in which a signal is on average higher for one group than another holding fixed productivity. When higher (lower) signal realizations lead to more favorable actions, this generates systemic discrimination against the group with the lower (higher) average signal. For example, Black defendants with the same potential for pretrial misconduct (Y_i^*) and underlying propensity for criminal activity (Y_i^0) as white defendants are less likely to have a clear criminal record (S_i), and this decreases the probability of being released on bail (Pager, Bonikowski, and Western 2009; Agan and Starr 2017).[19] Such signal inflation can arise from direct discrimination in another decision, e.g., Black individuals being more likely to be stopped by police and charged with a crime (Pierson, Simoiu, Overgoor, Corbett-Davies, Jenson, Shoemaker, Ramachandran, Barghouty, Phillips, Shroff et al. 2020).[20]

Signal inflation can be seen as statistical bias in the productivity signal. Thus, systemic discrimination from signal inflation can potentially be offset by an action rule that corrects for this bias, i.e., via "reverse" direct discrimination. Whether signal inflation translates to total discrimination depends on whether the manager is aware of the bias and chooses to correct for it. If she does not, then signal inflation will also lead to total discrimination. For example, suppose direct discrimination in policing leads to criminal record disparities, and hence, systemic discrimination in bail decisions. If a bail judge incorrectly believes that there is no direct discrimination in policing, and therefore, a prior criminal offense reflects the same underlying propensity for criminal activity across groups, then there will also be total discrimination in bail decisions. If instead the bail judge has accurate beliefs about discrimination in policing and accounts for it in her interpretation of criminal records, then her action rule will offset the systemic discrimination, resulting in no total discrimination. Importantly, however, it is possible that behavioral frictions can prevent information about

[19] The literature on systemic discrimination suggests many other examples of signal inflation. For instance, word-of-mouth recruitment practices that prioritize workers with certain social connections may lead to systemic discrimination when one group is more connected than equally qualified members of another group (perhaps because of past direct discrimination in hiring).

[20] Agan, Cowgill, and Gee (2021) show how signal inflation can potentially arise. In an audit study where job applicants disclose their prior salary, the distribution of disclosed salaries was chosen to match real-world gender differences, so that women on average reported lower salaries than men. They find little evidence for direct discrimination conditional on a given salary disclosure, but sizeable treatment disparities stemming from the lower disclosed salaries for women.

signal inflation from offsetting systemic discrimination. Section 5.2 and Appendix B.1 show how signal inflation leads to systemic and total discrimination, and that outcome bias can prevent a manager from correcting for signal inflation *even when* she is aware of the exact structure of bias in the signal-generating process. This highlights the utility of incorporating both direct and systemic discrimination within a common framework as it allows for the exploration of different policy interventions and the potential barriers to their efficacy.

Informational systemic discrimination can also arise when the manager has a more precise signal for group w than group b—what Cornell and Welch (1996) refer to as screening discrimination.[21] For example, suppose the signal is a test specifically trained to screen men and it generates less reliable information about the productivity of women.[22] De Plevitz (2007) similarly documents systemic discrimination due to the use of height-to-weight ratios calibrated with Anglo-Celtic data in job screening. Differential signal precision has also been noted as a source of statistical direct discrimination (Aigner and Cain 1977), which stems from the impact of the signal precision on the chosen action rule. In contrast, the systemic channel fixes the action rule and explores how differential signal precision translates to different action distributions. In Appendix B.2, we provide a theoretical and empirical example of how differential signal precision can lead to both direct and systemic discrimination. Together, these illustrate how canonical models of statistical discrimination fail to capture the full extent of screening discrimination.

Similar to the case of statistical direct discrimination (e.g., Fang and Moro 2011), systemic discrimination due to differential signal precision can be heterogeneous across qualification level. Consider, for example, a hiring decision in which the signal is equal to productivity plus mean-zero noise. A noisier signal hurts high productivity workers, as it leads to a higher chance of generating a signal below the hiring threshold, but can benefit low productivity workers by leading to a higher chance of a generating a signal above the hiring threshold. In contrast, in a medical diagnostic decision, all patients benefit from a more accurate signal when it leads to more accurate diagnoses regardless of health status. Notably, unlike signal inflation, systemic discrimination due to screening cannot be offset by the manager's action rule; to eliminate total discrimination, the manager needs to collect more precise information for group b (or ignore the more precise information for group w).

[21] In their model, minority job applicants receive fewer draws of a binary signal than majority applicants in a tournament setting where only one applicant is selected; as the number of applicants grows large, a majority applicant is hired with probability arbitrarily close to one. This is a form of systemic discrimination, given that were minority and majority applicants to receive the same number of signal draws, they would be evaluated equally (i.e., there is no direct discrimination). It can be directly nested in our framework if we allow for complementarities in the action rules across workers.

[22] Mocanu (2022) show that subjective tests designed to screen men led to disparate outcomes for women; amending or replacing the tests with more objective evaluations mitigated disparities. Pinkston (2003) also finds evidence that employers receive less-accurate initial signals from women than from men.

Technological Systemic Discrimination emerges from group-based differences in productivity for workers with the same qualification level. This channel is clearly only relevant when the chosen qualification is not productivity (i.e., $Y_i^0 \neq Y_i^*$). Similar to informational systemic discrimination, it can take the form of inflated productivity in which Y_i^* is systematically higher for group w workers than group b workers with the same qualification level. For example, white workers have more access to training and skill development than Black workers due to discrimination in education and the labor market.[23] Or, Black workers could anticipate future direct discrimination and respond by investing less in human capital (Coate and Loury 1993).[24] Technological systemic discrimination also includes the type of "task-based" discrimination studied in Hurst, Rubinstein, and Shimizu (2021), where workers have no initial group-based differences in qualification (Y_i^0) but racial barriers to specialization generate group-based differences in which tasks a worker chooses to specialize in (Y_i^*).

Discussion. Group-based differences in the distribution of worker qualification cannot lead to systemic discrimination with respect to that qualification by definition. This highlights how the chosen qualification reference impacts the potential sources of systemic discrimination. At the one extreme, when the qualification is set to a constant, all differences in the signal and productivity distributions contribute to systemic discrimination. At the other extreme, when the qualification is productivity, all systemic discrimination is informational. In between these two extremes, informational and technological channels can both contribute to systemic discrimination.

Statistical (direct) discrimination also stems from group-based differences in the signal and productivity distributions, but it conceptually differs from the sources of systemic discrimination we outline. Such statistical discrimination arises from the impact of the signal and productivity distributions on the action rule; in contrast, systemic discrimination arises from the impact of these distributions on the action *distribution* for a given qualification level. When $Y_i^0 \neq S_i$, differences in the signal distribution can lead to both informational systemic discrimination and accurate statistical direct discrimination. Similarly, when $Y_i^0 \notin \{Y_i^*, S_i\}$, differences in the productivity distribution can lead to technological systemic discrimination and accurate statistical direct discrimination. In both cases, focusing only on direct dis-

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

[24] In Coate and Loury (1993), workers have no group-based differences in initial productivity (Y_i^0). They make a costly decision to invest in human capital that increases productivity (Y_i^*). If Black workers believe they will face direct discrimination in hiring, they are less likely to invest in human capital. This supports the manager's equilibrium belief that Black workers have lower productivity, and Black workers face accurate statistical direct discrimination. The difference in the productivity distributions also results in group-based differences in the signal distributions for workers with the same qualification level, leading to systemic discrimination against Black workers.

crimination would miss a key aspect of how group differences in the signal and productivity distributions contribute to action disparities.

Finally, we note that unlike with statistical discrimination (e.g. Bordalo, Coffman, Gennaioli, and Shleifer 2019; Bohren, Haggag, Imas, and Pope 2022) there is no scope for "inaccurate" systemic discrimination: only the true productivity and signal distributions contribute to systemic discrimination through their statistical relationship with the qualification. However, inaccurate beliefs about the signal and productivity distributions can lead to inaccurate perceptions about the extent of systemic discrimination. This can affect action choices, and hence, total discrimination. It can also impact the choice of signaling technology. For example, a mortgage assessor may incorrectly believe that a particular credit score provides an identical signal of creditworthiness across groups, and therefore continue using it without adjusting for discriminatory signal inflation.

4 Measuring Systemic Discrimination

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

4.1 Preliminaries

We focus here on measures of discrimination that correspond to mean differences by group.[25] From Definition 2, a measure of total discrimination at qualification level $y^0 \in \mathcal{Y}^0$ is:

$$\Delta(y^0) \equiv E[A(G_i, S_i) \mid G_i = w, Y_i^0 = y^0] - E[A(G_i, S_i) \mid G_i = b, Y_i^0 = y^0]. \tag{3}$$

The first term corresponds to the expected action with respect to $\mu^w(a; y^0)$: the action distribution for group w at qualification y^0. The second term is with respect to $\mu^b(a; y^0)$: the action distribution for group b at y^0. For example, in a hiring decision with $A = 1$ for hire and $A = 0$ for do not hire, a finding of $\Delta(y^0) > 0$ means that group-w workers with qualification y^0 are more likely to be hired than equally-qualified group-b workers.

Correspondingly, from Definition 1, a measure of direct discrimination at signal realization $s \in \mathcal{S}$ is given by the difference between the selected actions for group w and group b when this signal is observed:

$$\tau(s) \equiv A(w, s) - A(b, s). \tag{4}$$

For example, in a hiring decision as described above, a finding of $\tau(s) > 0$ means that

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

belonging to group w versus b causes workers with signal s to be hired more often. A measure of the average level of direct discrimination at qualification level $y^0 \in \mathcal{Y}^0$ is given by:

$$\overline{\tau}(g, y^0) \equiv E[\tau(S_i) \mid G_i = g, Y_i^0 = y^0], \tag{5}$$

for $g \in \{w, b\}$. This *average direct discrimination* is the expected direct discrimination with respect to the signal distribution for group g at qualification y^0.

Finally, from Definition 3, a measure of systemic discrimination at qualification level $y^0 \in \mathcal{Y}^0$ is given by:

$$\delta(g, y^0) \equiv E[A(g, S_i) \mid G_i = w, Y_i^0 = y^0] - E[A(g, S_i) \mid G_i = b, Y_i^0 = y^0], \tag{6}$$

for $g \in \{w, b\}$. When $g = w$, this expression fixes the action rule for group w; the first term then corresponds to the expected action with respect to $\mu^w(a; y^0)$ and the second term is with respect to $\tilde{\mu}^b(a; y^0)$, the counterfactual action distribution for group b at y^0. Analogously, when $g = b$, the first term is with respect to counterfactual distribution $\tilde{\mu}^w(a; y^0)$ and the second term is with respect to $\mu^b(a; y^0)$. For example, in a hiring decision as described above, a finding of $\delta(g, y) > 0$ indicates higher hiring rates among equally-productive group-w versus group-b workers that arise indirectly from the signal.

While each of these measures are for a particular qualification level, it is also possible to construct an overall measure by averaging across qualification levels, as we do in Section 5. The interpretation of this overall measure depends on the chosen qualification distribution: averaging across the population qualification distribution yields a measure of average discrimination across both groups, while averaging across the qualification distribution for group g yields a measure of average discrimination for a group g worker.

4.2 Decomposing Total Discrimination

Our decomposition of total discrimination into direct and systemic components follows from Equations (3), (4) and (6), by adding and subtracting $E[A(b, S_i) \mid G_i = w, Y_i^0 = y^0]$ to and from the definition of $\Delta(y^0)$ and rearranging terms:[26]

$$\underbrace{\Delta(y^0)}_{\text{Total discrimination}} = \underbrace{\overline{\tau}(w, y^0)}_{\text{Avg. direct discrimination}} + \underbrace{\delta(b, y^0)}_{\text{Systemic discrimination}}. \tag{7}$$

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

Equation (7) shows that total discrimination at qualification level y^0 can be written as the sum of two terms: average direct discrimination with respect to the signal distribution for group w workers with qualification level y^0 and systemic discrimination at qualification level y^0 when the manager uses the action rule for group b.

Equation (7) is in the spirit of Kitagawa (1955), Oaxaca (1973), and Blinder (1973), who relate unconditional disparities to a component explained by observable worker characteristics (e.g., education or labor market experience) and a residual "unexplained" disparity. These classic decompositions can be viewed as a strategy for measuring direct discrimination, which attempts to hold fixed all relevant non-group characteristics. Equation (7), in contrast, leads to strategies (developed below) for measuring systemic discrimination as the residual of total discrimination after accounting for direct discrimination.

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

$$\Delta(y^0) = \overline{\tau}(b, y^0) + \delta(w, y^0) \tag{8}$$

by adding and subtracting $E[A(w, S_i) \mid G_i = b, Y_i^0 = y^0]$ to and from the definition of $\Delta(y^0)$ and rearranging terms. Equation (8) decomposes total discrimination into average direct discrimination with respect to the signal distribution for workers from group b and systemic discrimination when the firm uses the action rule for group w, all at qualification level y^0. Averaging Equations (7) and (8) yields a third decomposition:

$$\Delta(y^0) = \overline{\tau}(y^0) + \bar{\delta}(y^0), \tag{9}$$

where $\bar{\delta}(y^0) \equiv \frac{1}{2}(\delta(w, y^0) + \delta(b, y^0))$ averages the systemic discrimination terms and, slightly abusing notation, $\overline{\tau}(y^0) \equiv \frac{1}{2}(\overline{\tau}(w, y^0) + \overline{\tau}(b, y^0))$ averages the direct discrimination terms.

Each of the three decompositions (7)-(9) yield a measure of systemic discrimination, given by the difference between total discrimination and the direct discrimination component. The challenge of identifying systemic discrimination thus reduces to the challenge of measuring direct and total discrimination. We next discuss different identification strategies.

4.3 Observable Y_i^0: The Iterated Audit Design

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

When Y_i^0 is observed, a simple experimental approach can identify the direct and total discrimination components in equations (7)-(9). We term this approach an *iterated audit* (IA), as it applies tools from conventional audit or correspondence studies in multiple stages to empirically separate direct and systemic discrimination.[27]

The first IA step randomizes manager perceptions of group membership, as in a conventional audit or correspondence study, among a real set of workers with a given qualification level. Formally, in a population of workers with a given distribution of (G_i, S_i, Y_i^*, Y_i^0), the researcher generates a $\tilde{G}_i$ such that manager actions are given by $A_i = A(\tilde{G}_i, S_i)$ and where $\tilde{G}_i \perp\!\!\!\perp (G_i, S_i, Y_i^*) \mid Y_i^0$ by virtue of the randomization. This $\tilde{G}_i$ "treatment" can be used to measure the causal effect of group identity. For example, a researcher may take a set of real white and Black resumes and randomize distinctively white and Black names among equally-qualified workers, holding fixed all other information on the resumes.[28] The researcher then elicits manager actions A_i in the experimental sample. Comparing the response to group-w resumes randomized to $\tilde{G}_i = w$ with the response to group-w resumes randomized to $\tilde{G}_i = b$, at qualification level y^0, identifies the direct discrimination component of (7):

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

Similarly, comparing the response to group-b resumes randomized to $\tilde{G}_i = w$ with the response to group-b resumes randomized to $\tilde{G}_i = b$, at qualification level y^0, identifies the direct discrimination component of (8). Averaging these comparisons identifies the direct discrimination component of (9).

The second IA step measures *total discrimination* by eliciting manager actions among workers whose perceived group membership was not manipulated by the experiment. This could be in a separate non-experimental sample, with the same distribution of (G_i, S_i, Y_i^*, Y_i^0), or among workers with $G_i = \tilde{G}_i$ in the experimental sample. Subtracting one of the three direct discrimination components estimated in the first step from the total discrimination measure $\Delta(y^0) = E[A_i \mid G_i = w, Y_i^0 = y^0] - E[A_i \mid G_i = b, Y_i^0 = y^0]$ identifies one of the three systemic discrimination components in equations (7)-(9).

Figure 1 illustrates an example iterated audit conducted with white (group w) and Black (group b) resumes, where the researcher is interested in studying discrimination conditional on worker education Y_i^0. Resumes A and C represent "endogenous" profiles of white and

[27]The IA method also applies when Y_i^0 is unobserved but independent of G_i and not a component of S_i.

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

Figure 1. Iterated Audit Example

Black applicants (i.e., the resume from an actual white worker, $G_i = w$, that was assigned to a perceived race of white, $\tilde{G}_i = w$) with the same level of education (i.e., same y^0). Disparities in hiring decisions (such as callback rates) between such resumes capture total discrimination for equally-educated workers. Resume B represents an "exogenous" profile of a Black applicant with all observable information matching the white candidate's; this resume is generated by randomizing a distinctively Black photo ($\tilde{G}_i = b$) to the real white resume A ($G_i = w$). Comparing A and B implicitly holds fixed all non-race elements, such as education and work experience; hence, hiring disparities arising from this randomized "treatment" capture direct discrimination as in a classic correspondence study. Finally, comparing resumes B and C captures systemic discrimination—hiring disparities due to non-group characteristics among equally-qualified workers perceived to be of the same group.

Outside of experimental settings, the core IA logic can be applied to the observable Y_i^0 case whenever direct discrimination can be reliably measured. For example, if manager signals S_i are observed by the researcher, then direct discrimination is identified from $\tau(s) = E[A_i \mid G_i = w, S_i = s] - E[A_i \mid G_i = b, S_i = s]$. Average direct discrimination can be constructed from these and the conditional distribution of S_i given (G_i, Y_i^0). Subtracting the measure of average direct discrimination from the identified total discrimination measure again yields a corresponding measure of systemic discrimination from Eq. (7)-(9).[29]

4.4 Selectively Observed or Proxied Y_i^0

In some cases, the researcher-chosen measure of qualification may be only selectively observed given the manager's actions. For example, when $Y_i^0 = Y_i^*$ measures a worker's productivity in the task at hand and $A_i \in \{0, 1\}$ indicates a hiring decision, observed output $Y_i = A_i Y_i^0$ gives a selective measure of qualification: workers who are hired ($A_i = 1$) reveal their

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

productivity but Y_i^0 is unobserved among unhired workers. Selective observability may also pose a challenge when Y_i^0 is an "upstream" measure of productivity, such as when a worker first enters the labor market prior to the current task.

The IA approach to measuring systemic discrimination can be adapted to this case by incorporating additional (quasi-)experimental variation to address the new selection challenge. Arnold et al. (2022), for example, develops quasi-experimental methods to study disparate impact in pretrial release decisions by leveraging the as-good-as-random assignment of individuals to bail judges. They show how such assignment can be used to "selection-correct" the observed distribution of qualification by group, and how the resulting unselected qualification distribution can be used to estimate total discrimination by a particular adjustment of the unconditional group disparities $E[A_i \mid G_i = w] - E[A_i \mid G_i = b]$.[30]

To translate their approach to the hiring example, suppose managers with potentially different hiring rates are as-good-as-randomly assigned to workers.[31] Combining experimental variation in group membership perceptions—as in a classic audit or correspondence study—with the (quasi-)experimental action variation underlying the Arnold et al. (2022) approach yields a measure of systemic discrimination. Specifically, consider a set of group-w workers with experimentally manipulated group perceptions among as-good-as-randomly assigned managers. The direct discrimination component in Eq. (7) can be estimated in this subsample by using the quasi-experimental selection correction technique to adjust the experimental disparities $E[A_i \mid \tilde{G}_i = b, G_i = w] - E[A_i \mid \tilde{G}_i = w, G_i = w]$. Subtracting this term from the Arnold et al. (2022) measure of total discrimination identifies the systemic discrimination component in Eq. (7). Analogous steps identify the other decompositions.

A harder identification challenge arises when Y_i^0 is not even selectively observed and must be proxied by other observables X_i. Here the frameworks of Altonji et al. (2005) and Oster (2019) may be integrated in the IA approach to measure systemic discrimination. Specifically, one can use these frameworks to bound or estimate total discrimination from unconditional and conditional-on-X_i disparities by making assumptions about how the effect of conditioning on X_i relates to the effect of the infeasible conditioning on Y_i^0. In samples where group membership is experimentally manipulated, such extrapolations may be further used to bound or estimate average direct discrimination—and therefore systemic discrimination.

To summarize, the IA approach can be used to estimate each of the three decompositions (7)-(9) by leveraging a combination of (quasi-)experimental and observational variation. We note that different choices of the qualification metric Y_i^0 may require different sources of

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

[31] In practice, manager assignment can be substituted with any (quasi-)experimental variation in actions that allows for such correction of the selected observed qualification distribution.

variation and identification strategies. Moreover, with multiple choices of Y_i^0 it is possible to further decompose total discrimination into a direct component and multiple systemic components, e.g., to reflect different informational or technological sources. Bringing these richer decompositions to data follows similarly as above.

5 Applications

This section presents three empirical applications. The first, in Section 5.1, illustrates the IA methodology and shows how direct discrimination in a prior decision can lead to sizable systemic discrimination in subsequent actions. The second, in Section 5.2, shows how information-processing frictions such as outcome bias can blunt the effectiveness of informational interventions meant to eliminate systemic discrimination. It also illustrates the importance of the reference qualification, showing how different choices can meaningfully change measures of total and systemic discrimination. The third, in Section 5.3, shows how direct discrimination in one decision can generate technological systemic discrimination in a contemporaneous decision, highlighting that systemic discrimination does not only arise from sequential decision-making. We discuss two additional applications in Appendix B that illustrate how our framework can be used to measure systemic discrimination from signal inflation and screening.

5.1 Iterated Audit

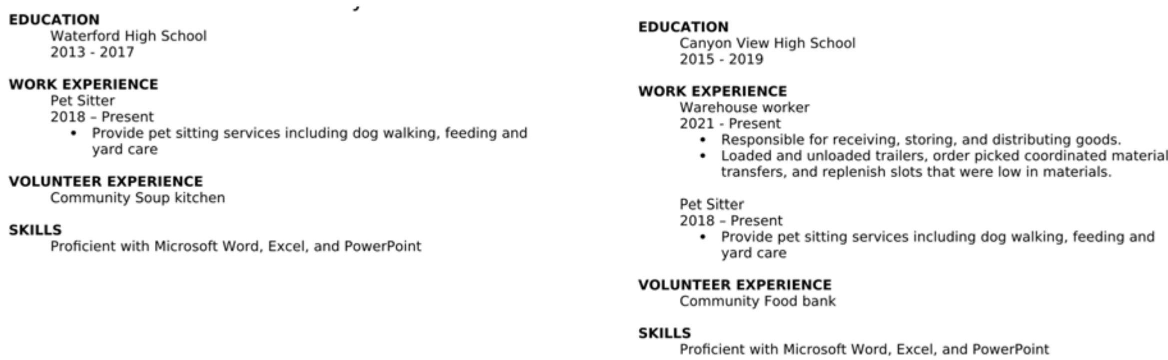
We employ a lab-in-the-field experiment to show how systemic discrimination arises in sequential hiring decisions and to illustrate the IA methodology from Section 4. Specifically, we show that past direct discrimination in hiring leads to systemic discrimination in subsequent hiring—what Feagin and Feagin (1978) call "past-in-present" discrimination.

5.1.1 Setup

We used a hiring and recruitment agency to recruit hiring managers ($N = 208$) with experience in evaluating applicants to entry-level jobs and who were currently looking for employees. Hiring managers evaluated fictitious resumes to an entry-level job on the likelihood of the applicant being hired for the job on a scale of 1 to 10. Decisions were incentivized using a similar methodology to Kessler et al. (2019): the resumes themselves were fictitious, but the components (e.g., prior work experience) could be matched to resumes of actual potential applicants who had similar attributes and presented to the managers based on their likelihood scores.[32]

Our IA design featured three sets of resumes, as depicted in Figure 1. Two of the

[32]This factorial design is known as an Incentivized Resume Rating paradigm. See Lahey and Oxley (2021) and Kübler, Schmid, and Stüber (2018) for similar uses of factorial designs in studying discrimination.



EDUCATION
Waterford High School
2013 - 2017

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

VOLUNTEER EXPERIENCE
Community Soup kitchen

SKILLS
Proficient with Microsoft Word, Excel, and PowerPoint

EDUCATION
Canyon View High School
2015 - 2019

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

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

VOLUNTEER EXPERIENCE
Community Food bank

SKILLS
Proficient with Microsoft Word, Excel, and PowerPoint

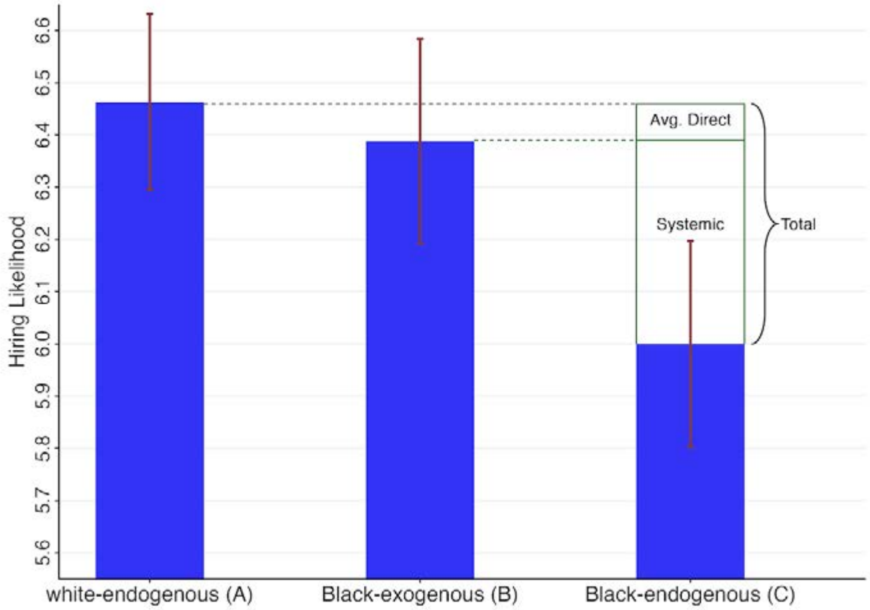
(a) No Entry-Level Work Experience

(b) Entry-Level Work Experience

FIGURE 2. Iterated Audit: Example Resumes

three sets, referred to as white-endogenous (A) and Black-exogenous (B), were similar to those in a standard correspondence or audit study: the resumes were the same aside from perceived group identity (in our case, distinctively white or Black name). These resumes were constructed by assigning work experience based on a "white" distribution. The third set of resumes, Black-endogenous (C), had Black perceived group identity (distinctively Black names) but differed from set (B) in that work experience was assigned based on a "Black" distribution. The distributions of "white" and "Black" work experience were constructed based on a previous audit study by Pager (2003). There, matched pairs of individuals applied for an entry-level job. Black applicants were found to be significantly less likely to proceed through the application process than white applicants with the same qualifications (14% versus 34%). We use these differential rates to assign entry-level work experience to each set of resumes. Specifically, we generated six different resumes in each set, A, B and C, that differed in name and work experience. In the white-endogenous (A) and Black-exogenous (B) sets, 33% of resumes (2 out of 6) had entry-level work experience based on the "white" rate from Pager (2003), and the remaining resumes had no entry-level work experience. In contrast, in the Black-endogenous set (C), 17% of resumes (1 out of 6) had entry-level work experience based on the "Black" rate from Pager (2003). See Figure 2 for examples.

Each hiring manager evaluated a random draw of four resumes. The managers were told about the incentive structure and the results of the Pager (2003) study—particularly the differential rates that white versus Black applicants were contacted to proceed with the entry-level job. Managers were thus informed about potential racial disparities in signals (i.e., entry-level work experience). This design choice was used to generate a conservative test for how systemic factors impact total discrimination in this setting.



Note: Whiskers indicate +/- one manager-clustered standard error.

FIGURE 3. Iterated Audit: Hiring Likelihoods

5.1.2 Results

We set the qualification as the content of the resume prior to the addition of entry-level work experience (or lack thereof).[33] All other components of the resume (e.g., education, volunteer experience, skills) were randomly assigned and chosen to be similar across workers. In turn, workers in each set have the same qualification distribution. We measure discrimination in terms of differences in the average hiring likelihood, averaged across the qualification.

This IA design allows us to identify the measures of total, systemic and direct discrimination in Eq. (7) by a simple comparison of means.[34] Figure 3 presents this breakdown graphically while Table 1 presents the results in regression form. Total discrimination corresponds to the difference in hiring likelihood between the white-endogenous and Black-endogenous resumes (A vs. C). We find significant total discrimination: there was a hiring likelihood gap of 0.46 between these two groups (Column 1 of Table 1), corresponding to 19% of a standard deviation in hiring likelihoods. Comparing white-endogenous and Black-exogenous resumes (A vs. B)—the standard comparison in audit and correspondence studies—provides a measure of average direct discrimination. While the coefficient on race is negative (Column 2 of Table 1), it is small and statistically insignificant; as shown in Figure 3, it explains only a

[33]In terms of our model, this corresponds to taking the signal vector S_i that contains all relevant components of the resume and, letting $S_{i,k}$ denote the component that documents entry-level work experience, removing this component, $Y_i^0 = S_i \setminus \{S_{i,k}\}$.

[34]Alternative choices for group B would similarly identify the measures in Eqs. (8) and (9).

Table 1. Iterated Audit: Hiring Likelihood Regressions

	Total (1)	Avg. Direct (2)	Systemic (3)
Black Resume Dummy	-0.463**	-0.075	
	(0.201)	(0.208)	
Endogenous Resume Dummy			-0.388*
			(0.216)
Included Resume Sets	A,C	A,B	B,C
# Observations	583	535	548

Notes: This table reports coefficients from regressing hiring likelihoods on dummies for resume group. Manager-clustered standard errors are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

small portion of total discrimination in our setting. On the other hand, Figure 3 shows that systemic discrimination—the difference between the Black-exogenous and Black-endogenous resumes (B vs. C)—drives the vast majority of total discrimination (Column 3 of Table 1). This is consistent with total discrimination being driven in large part by the upstream direct discrimination that generated disparities in work experience. Note that we cannot identify whether this systemic discrimination has a technological source, informational source (in the form of signal inflation), or a combination of the two. This depends on whether the hiring managers believed that the entry-level work experience increased a worker's productivity or simply signaled information about productivity.

Strikingly, the prior discrimination in hiring impacts current hiring despite the managers being told that work experience disparities were likely generated by direct discrimination elsewhere in the system. While this information may have reduced the extent of direct discrimination (per Column 2 of Table 1), systemic discrimination still led to substantial differences in hiring likelihoods. In light of prior work showing the effectiveness of information about group-based differences in productivity in reducing direct discrimination (Bohren et al. 2022), these findings highlight the difficulty of using a similar tactic to mitigate total discrimination when it is caused by systemic factors.

5.2 Mitigating the Impact of Systemic Discrimination

The previous study provided a conservative measure of how systemic factors impact total discrimination by informing evaluators that direct discrimination had occurred at earlier stages. As outlined in Section 3.4, awareness of prior direct discrimination can potentially mitigate the impact of systemic discrimination that stems from signal inflation, as it allows evaluators to correct for the bias in the signal. At the same time, information about systemic factors may fail to mitigate disparities due to psychological frictions such as outcome bias (Baron and Hershey 1988; Brownback and Kuhn 2019). Its effectiveness may also be blunted by evaluators' "home grown" preferences and beliefs that they import from outside of the setting, uncertainty about the data-generating process, or distrust about the information provided.[35] We designed our next study to hone in on outcome bias in perpetuating the impact of systemic discrimination. We employed an abstract setting with a transparent data-generating process, clear incentive structure, no scope for "home grown" beliefs or preference against groups in the study, and no productivity differences between groups. We also use this design to demonstrate how the measurement and interpretation of discrimination depend critically on the chosen qualification.

5.2.1 Setup

The experiment was conducted on the *Prolific.co* platform. All participants were paid a \$1.50 base fee for completing the study.[36]

Workers: 100 participants were randomly assigned to the role of Worker. Each Worker completed a series of exercises in which they counted the number of zeroes in a matrix table containing zeroes and ones. Workers were given five minutes to complete up to ten of the tables.[37] Worker i's productivity Y_i^* is the number of correctly solved tables. Workers were then assessed by two sets of evaluators

Recommenders: 300 participants were randomly assigned to be the first set of evaluators, termed Recommenders. Recommenders were presented with profiles of five Workers and incentivized to predict each Worker's productivity based on a noisy signal.[38] The signal was either biased upward or downward depending on the Worker's group membership. Here, we follow the work of Esponda, Oprea, and Yuksel (2023) and Dianat, Echenique, and Yariv (2022) in randomly assigning group membership as one of two abstract categories, $G_i \in \{a, b\}$, with equal probability. This ensures ex-ante neutral beliefs and preferences toward members of each group. Signals from Workers assigned to Group a had upwardly-biased noise added to productivity, while signals from those assigned to Group b had downwardly-

[35] Note that when systemic discrimination has a technological source, systemic factors impact productivity directly and information about such factors is not necessarily predicted to undo them.

[36] See the Online Appendix for experimental instructions.

[37] Each table took about one minute to complete. Workers were paid an additional \$0.20 for every table completed correctly.

[38] Specifically, Recommenders were told that one of the five Worker profiles would be picked at random and they would earn a bonus of \$2 if they correctly predicted that Worker's productivity. We used this mechanism to incentivize beliefs based on the work of Danz, Vesterlund, and Wilson (2022), who show that more complicated mechanisms such as the binarized scoring rule lead to conservatism in elicited beliefs and greater error rates compared to simpler mechanisms. They argue that incentives based on belief quantiles—such as the one we use here—will result in more truthful reporting and lower cognitive burden.

biased noise. Specifically, $S_i = Y_i^* + e_i^{G_i}$, where $e_i^a \sim N(1,2)$ and $e_i^b \sim N(-1,2)$.[39] In order to simplify the inference problem, we induced a uniform prior over Workers' productivity by selecting a group of Worker profiles from the full sample so that an equal number of Workers achieved each level of productivity between 1 and 7.[40] Recommenders were given full information about the Workers' task, how the sample was constructed, how group was assigned, and how signals were generated based on group membership; the latter was presented both visually and in words. Each Recommender then had to pass a series of comprehension checks that tested their understanding of the bias.

Recommenders were randomly assigned to one of two treatments. In the Informed treatment, they also observed each Worker's group and could, in theory, fully adjust for the bias in the signal; in the Group-Blind treatment, they had the same information about the signal-generating process but did not observe each Worker's group. After observing a Worker's signal realization and potentially the Worker's group, the Recommender submitted an evaluation (prediction of productivity) between 1 and 7.

Given the transparency of the signal bias and the incentives to correctly guess each Worker's underlying productivity, the Informed treatment is a conservative test for the efficacy of information about signal bias to correct for systemic discrimination. Similar real-world interventions would likely be blunted by the other frictions discussed above.

Evaluators: 400 participants were randomly assigned to be the second set of evaluators, termed Evaluators. Evaluators were provided with the same description of the environment as Recommenders and similarly incentivized to predict the productivity of five Workers. However, instead of observing a noisy signal of productivity, they observed an Recommender's assessment of that Worker's productivity. They were matched with Recommenders from the Group-Blind treatment and informed that the Recommender did not observe the Workers' group. Evaluators were randomly assigned to one of two treatments that mirrored the Recommenders' two treatments. In the Informed treatment, Evaluators observed each Worker's group and could therefore adjust for potential bias in Recommenders' evaluations; in the Group-Blind treatment, Evaluators did not observe each Worker's group.

We measure discrimination with respect to differences in evaluations for Groups a and b. Disparities in evaluations can arise because Group a's evaluations are upwardly biased relative to actual productivity, Group b's evaluations are downwardly biased, or both. We present an analysis of this evaluation bias in Appendix C.1. Recall that Workers were randomly assigned to a group, group names were abstract, and there were no group-based differences

[39] Using a normal distribution to generate noise, as opposed to a distribution with a finite support, ensures that no signal realization is perfectly diagnostic of productivity.

[40] Coffman, Kostyshak, and Saygin (2023) used a similar methodology in an experiment studying gender discrimination in information acquisition.

in the distributions of productivity; hence any disparities are unlikely to be driven by direct discrimination due to preferences or beliefs about group differences in productivity.

5.2.2 Results

This design has two key features: it allows us to illustrate how the measurement of discrimination depends critically on the chosen qualification, and it isolates the role of outcome bias in perpetuating systemic discrimination. We consider two choices for the reference qualification: the information observed by the evaluators, e.g., $Y_i^0 = S_i$ in the case of Recommenders, which focuses on direct discrimination, and the underlying productivity, $Y_i^0 = Y_i^*$, which accounts for informational systemic factors. Additionally, we compare how discrimination responds to information about exogenous signal biases (Recommenders) versus endogenous signal biases stemming from others' evaluations (Evaluators). The latter is more complex to correct as it requires an understanding of how others interpret information.

Recommenders: Table 2 presents results of our decomposition that separately identifies total, direct, and systemic discrimination based on the chosen Y_i^0. In the Group-Blind treatment, total discrimination is negative and significant when productivity is the reference qualification, $Y_i^0 = Y_i^*$: on average, Group b workers receive 0.85 points (0.59 standard deviations of the mean) lower evaluations than Group a workers (Column 1 of Table 2). Nearly all of this total discrimination is due to systemic factors: systemic discrimination lowered the evaluations of Group b workers by 0.81 points, compared to a small and insignificant estimate of average direct discrimination (0.04). If we instead choose the signal as the reference qualification, $Y_i^0 = S_i$, we reach the starkly different conclusion: the small estimate of 0.01 suggests that there is little to no total (or direct) discrimination.[41]

Informed Recommenders were aware of the exact structure of the signal bias and could therefore (in theory) correct for it. Total discrimination remained negative and substantial with productivity as the reference qualification, but was smaller in magnitude compared to Group-Blind Recommenders: on average, Group b workers receive 0.46 points (0.26 standard deviations) lower evaluations than Group a workers (Column 2 of Table 2). Systemic factors generated an even larger disparity: systemic discrimination lowered the evaluations of Group b workers by 0.53 points, but this was partially offset by "reverse" direct discrimination that favored Group b by 0.07 points.[42] Thus, awareness of bias in the signal-generating

[41]To estimate average total discrimination across a continuous qualification (the signal), we ran a linear regression of evaluation on group, the signal, and their interaction. The linear approximation yields an estimate of direct discrimination for each worker at her realized signal. We approximate average direct (and total) discrimination by taking an average of all workers' estimated direct discrimination. See Table 9 in Appendix C.1 for the intermediate regression results.

[42]Note that systemic discrimination need not be the same between the two treatments, since it is a function of both the (same) signal distribution and the (potentially different) action rule.

Table 2. Mitigation Study: Discrimination Decomposition

	Recommender		Evaluator	
	Group-Blind (1)	Informed (2)	Group-Blind (3)	Informed (4)
Qualification: Productivity				
Total: $E_{y^*}[\Delta(y^*)]$	-0.85***	-0.46***	-0.95***	-1.04***
	(0.09)	(0.12)	(0.08)	(0.11)
Average Direct: $E_{y^*}[\overline{\tau}(y^*)]$	-0.04	0.07***	-0.01	0.11
	(0.07)	(0.03)	(0.07)	(0.12)
Systemic: $E_{y^*}[\overline{\delta}(y^*)]$	-0.81***	-0.53***	-0.94***	-1.16***
	(0.10)	(0.13)	(0.07)	(0.09)
Qualification: Signal				
Total (=Direct): $E_s[\Delta(s)]$	0.01***	0.66***	-0.02***	0.08***
	(0.00)	(0.01)	(0.00)	(0.00)
# Observations	780	720	1055	945

Notes: This table reports estimates of each measure of discrimination in Eq. (9), averaged across the qualification (i.e. either by the population distribution of productivity or the population distribution of the signal). See Appendix C for the other two versions of the decomposition. The sample includes 300 Recommenders and 400 Evaluators, each evaluating five Workers. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

process resulted in Recommenders correcting for it, but only partially—the majority of total discrimination remained.

Choosing the signal as the reference qualification, $Y_i^0 = S_i$, again yields a starkly different conclusion: namely, that the information about bias led Recommenders to favor Group b workers and discriminate against Group a workers instead. Total discrimination was positive, sizeable, and significant at 0.66, indicating substantial "reverse" total (and direct) discrimination. Importantly, however, concluding that information about signal bias leads to reverse discrimination is misleading: rather than giving Group b an edge, reverse direct discrimination offset the systemic factors that disadvantage Group b in the first place.

Taken together, these results show that Recommenders do not fully account for signal bias despite being aware of its exact structure and having clear incentives to do so—the majority of total discrimination remains. This highlights outcome bias as an important friction in individual-level informational interventions as a strategy to mitigate the impact of systemic discrimination.

Evaluators: Recommenders observed an exogenous signal generated by a transparent process. Evaluators, on the other hand, observed an endogenous productivity signal: the Recommender's evaluation. Here the problem is more complex because inference requires an understanding of how Recommenders interpreted their information. We next show that processing this endogenous signal exacerbated the impact of outcome bias.

Group-Blind Evaluators mirror the case of Group-Blind Recommenders (Column 3 of Table 2). Total discrimination is -0.95 when fixing productivity as the qualification, $Y_i^0 = Y_i^*$, which is significant, sizeable, and of similar magnitude to Recommenders. Again, this is almost entirely driven by systemic factors. When fixing the signal as the qualification (in the case of Evaluators, S_i is the Recommender's evaluation), total discrimination falls to a negligible level (-0.02). Therefore, both exogenous and endogenous bias in the signal-generating process led to substantial total discrimination in the Group-Blind condition, which was driven by systemic factors and would be missed if the signal was chosen as qualification.

Informed Evaluators observed the Workers' groups but were aware that the Recommenders did not.[43] Hence, Informed Evaluators could potentially parse out the signal bias. Indeed, we find that Informed Evaluators' patterns of discrimination are qualitatively different from those of Informed Recommenders. When holding fixed productivity as the qualification, $Y_i^0 = Y_i^*$, total discrimination remains large, significant, and of similar magnitude to the Group-Blind Evaluators: evaluations of Group b workers are 1.04 points lower than Group a workers with the same productivity (Column 4 of Table 2). This contrasts with Informed Recommenders, who exhibited smaller levels of total discrimination as Group-Blind Recommenders (Columns 1 and 2 of Table 2)—though more than half remained. In other words, more information about the signal process did not reduce total discrimination when the signal process was endogenous.

When fixing the signal as the qualification, Informed Evaluators exhibited a small level of direct/total discrimination (Column 4 of Table 2). This contrasts with Informed Recommenders, who exhibited substantial reverse direct discrimination that favored Group b. Thus, when faced with an endogenous signal, there seemed to be a smaller correction for systemic factors. One plausible explanation for this is *information projection* (Madarász 2012), where Evaluators failed to internalize that Recommenders did not also observe Worker group—and hence, did not already correct the signal bias—when forming their evaluations.

Together, these results suggest that a biased endogenous signal process may be even more difficult to correct via an individual-level informational intervention, as outcome bias may interact with other biases (e.g., information projection) and other factors, such as uncertainty about how to account for discrimination in the evaluations of others. This supports the view that the transparent signal-generating process case is indeed an upper bound on the effectiveness of such an individual-level informational intervention. It also highlights the need for alternative interventions in mitigating the impact of systemic discrimination.

[43] Recall that Evaluators in the Informed treatment observed the evaluations of Recommenders from the Group-Blind Treatment.

5.3 Contemporaneous Discrimination

The previous two studies show how past decisions contribute to systemic and total discrimination. As discussed in Section 3.2, contemporaneous and future decisions can also contribute to both forms of discrimination. We next illustrate how direct discrimination by evaluators in one setting (e.g., by one company) can generate contemporaneous systemic discrimination by evaluators in another setting (e.g., by another company) when both sets of evaluators make decisions simultaneously (or more generally, when evaluators in neither setting observe the decisions of the other before making their own). Specifically, we show that when a worker makes an endogenous investment in productivity and this productivity is relevant in multiple settings, then anticipating direct discrimination in one setting reduces the worker's investment. This leads to technological systemic discrimination in the other setting, even though evaluators in this other setting do not themselves engage in direct discrimination. The results thus show that systemic discrimination need not emerge from dynamic decisions: it can also emerge from contemporaneous interactions—what Feagin and Feagin (1978) call "side-effect" discrimination.

5.3.1 Setup

We recruited 100 participants from *Prolific.co* to complete a series of tasks as described for Workers in the previous study. All participants were paid a \$3.20 base fee for completing the study. Participants were presented with a sample table and then chose how many of them to complete (0 to 10) in each of three treatments (described below). Participant i thus makes a productivity choice Y_i^* in a given treatment, which corresponds to the number of tasks she chooses to complete. In each treatment, the participant was randomly assigned to one of two groups, $G_i \in \{a, b\}$, and then evaluated by two companies.

Company 1 engaged in direct discrimination against Group b by construction. It observed the participant's group and a signal $S_{i,1} = Y_i^* + e_{i,1}$ of productivity, where $e_{i,1} \sim N(0, 2)$, then evaluated the participant according to action rule $A_1(a, S_{i,1}) = S_{i,1} + 3$ and $A_1(b, S_{i,1}) = S_{i,1} - 3$.[44] This implies direct discrimination against Group b, since, holding fixed the signal, Group a participants received systematically higher evaluations, $A_1(a, s) > A_1(b, s)$. Given evaluation $A_{i,1} = A_1(G_i, S_{i,1})$, Company 1 paid a bonus of $A_{i,1}/10$ dollars.

Company 2 did not engage in direct discrimination. It observed a signal $S_{i,2} = Y_i^* + e_{i,2}$ of productivity, where $e_{i,2} \sim N(0, 2)$, then evaluated the participant according to the group-blind action rule $A_2(g, S_{i,2}) = S_{i,2}$.[45] Given evaluation $A_{i,2} = A_2(G_i, S_{i,2})$, Company 2 paid a bonus of $A_{i,2}/10$ dollars.

[44] Thus, relative to productivity, Participant i received a positively biased evaluation $A_{i,1} = Y_i^* + e_{i,1} + 3$ if assigned to Group a and a negatively biased evaluation $A_{i,1} = Y_i^* + e_{i,1} - 3$ if assigned to Group b.

[45] Thus, Participant i received an unbiased evaluation $A_{i,2} = Y_i^* + e_{i,2}$ of productivity, regardless of group.

Participants completed three treatments: they chose how many tables to solve after being assigned to Group a, being assigned to Group b, and before learning group assignment. We focus on the first two treatments for our analysis.[46] One of these treatments was randomly selected, the participant completed her chosen number of tasks, and was then paid according to the process described above. The evaluation process and bonus payment procedure was described to participants before they chose the number of tables to complete; a series of comprehension questions checked that they understood this procedure.

We model the participant's productivity decision by considering each participant's *cost of productivity* $c_i : \mathcal{Y} \to \mathbb{R}_+$, which captures the effort cost of generating a given level of productivity; this cost captures factors such as the time and attention required to complete the task given the participant's current skill level and the opportunity cost of time. When choosing productivity, the participant trades-off this cost with the expected return to productivity, i.e., the bonus payment. Importantly, the within-subject design allows us to compare how chosen productivity varied with group assignment while holding fixed the participant's underlying cost of productivity.

5.3.2 Results

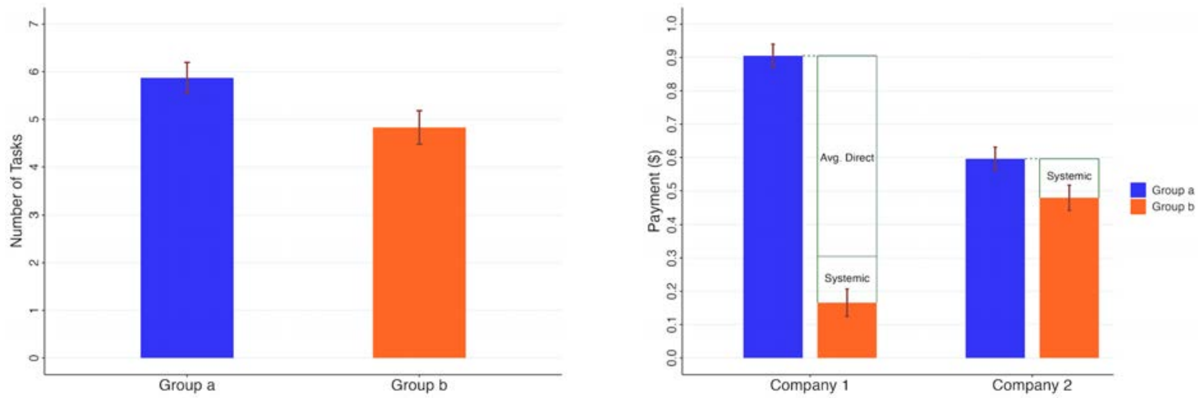
We choose the cost of productivity as the reference qualification, $Y_i^0 = c_i$, in order to hold fixed all factors that impact the productivity choice aside from the anticipated return to productivity. While this is not directly observable for each participant, the within-subject design ensures that its distribution is the same across groups. We measure discrimination in terms of the difference in bonus payments for Groups a and b, averaged across this qualification.

Figure 5(a) compares the average productivity choice Y_i^* for Group a and Group b participants. Relative to Group a participants, those in Group b selected a lower productivity in response to the anticipated direct discrimination by Company 1. Hence, the expected return to productivity does indeed impact participants' investment decision. This illustrates how anticipated disparities in the return to productivity can generate endogenous group-based differences in the distribution of productivity.

We next show that these endogenous productivity differences cause *both* companies to exhibit technological systemic discrimination in payments. Table 3 presents results from our decomposition that unpack total discrimination in payments from each company into its direct and systemic components; Figure 5(b) presents these results graphically.[47] Unsurprisingly, Company 1 paid Group a participants more than Group b participants: on average,

[46]We present the analysis of the group-blind treatment in Appendix C.2. The results are unsurprising and lie in between the two observe-group treatments.

[47]Since average direct discrimination is constant across the signal distribution for both companies, all three versions of the decomposition (i.e., Eqs. (7) to (9)) are equivalent.



(a) Productivity Choice: # Tables Completed

(b) Bonus Payment by Company

Note: Whiskers indicate +/- one robust standard error.

FIGURE 4. Contemporaneous Study: Productivity and Payments

Group a participants received \$0.91 and Group b participants received \$0.17, leading to total discrimination of \$0.74 averaged across qualification (cost of productivity). Part of this disparity was driven by direct discrimination: it follows from the action rule that, fixing the signal, Group a participants were paid \$0.60 more than Group b participants. However, nearly 20% of total discrimination (\$0.14) comes from Group b participants making lower productivity choices than Group a participants with the same reference qualification. This proportion of the gap corresponds to technological systemic discrimination, as it stems from the interaction between Company 1's evaluation decision and the productivity choice. It is technological because it arises from endogenous group-based productivity differences amongst participants with the same cost of productivity.[48]

Company 1's direct discrimination led to contemporaneous systemic discrimination at Company 2: despite not engaging in any direct discrimination, Company 2 paid Group b participants on average 20% (\$0.12) less than Group a participants. This disparity is driven by the lower productivity choices of Group b participants relative to Group a participants with the same reference qualification. Thus, as with Company 1, this disparity corresponds to systemic discrimination but now stems from the interaction between *three* decisions: the evaluation decisions of both companies and the productivity decision of the participant. Anticipation of direct discrimination in one evaluation drove a pre-evaluation productivity decision that then led to systemic discrimination in another contemporaneous evaluation. Since the latter did not engage in direct discrimination, the entirety of total discrimination—corresponding to the 20% gap in payments—was driven by systemic discrimination.

[48]Note that Group a participants were paid \$0.60 more than equally-productive Group b participants. Thus, total discrimination with respect to productivity is equal to average direct discrimination (0.6), and hence, there is no informational systemic discrimination between participants with the same productivity.

Table 3. Contemporaneous Study: Discrimination Decomposition

	Company 1	Company 2
Total	0.739***	0.117***
	(0.03)	(0.04)
Avgerage Direct	0.600	0.000
	(N/A)	(N/A)
Systemic	0.139***	0.117***
	(0.03)	(0.04)
# Observations	200	200

Notes: This table reports estimates of each measure of discrimination in Eqs. (7) to (9) (all three are equivalent), averaged across the cost of productivity as qualification. The direct numbers follow from the action rules and the observation that, given direct discrimination is constant with respect to the signal, average direct is also equal to 0.6 or 0, respectively; the total estimates are the difference between average bonus payments to Group a and Group b workers; the systemic estimates are equal to total minus average direct. The sample includes 100 participants, each making two decisions. Standard errors are reported in parenthesis; since the direct numbers are calculations not estimates, they do not have SEs. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

6 Conclusion

Large literatures, mostly from outside of economics, emphasize the importance of systemic factors in driving group-based disparities; yet economic analyses largely focus on direct discrimination as a function of of group identity itself. We bridge this gap by developing new theoretical and empirical tools to study systemic discrimination. Our general theoretical framework nests the canonical notion of direct discrimination with broader notions, and formalizes the importance of researchers taking one (or several) explicit stances on individual qualification for a given action. Our empirical decomposition of total discrimination into direct and systemic components further motivates the development of new econometric tools that identify these components in experimental and observational data. Our empirical applications, including the novel Iterated Audit design, show how conventional methods of studying direct discrimination can miss total discrimination and important heterogeneity in practice. These also show how frictions such as outcome bias can blunt the effectiveness of individual-level interventions in mitigating the impact of systemic factors.

By formalizing the differences and possible interactions between direct and systemic discrimination, our framework can be useful for interpreting and predicting the effects of policies aimed at reducing disparities. Consider the case of racial or gender quotas. In standard models of taste-based or statistical discrimination, such policies would have a temporary effect on disparities: evaluators' decisions would revert back and the disparity would re-emerge when a quota is lifted. However, if the initial disparity was due to technological systemic discrimination, e.g., in access to skill development, then quotas may reduce the disparity in the skill distribution as they create an incentive to develop female players. De Sousa and Niederle (2022) show that the introduction of a team quota for the minimum number of female chess players improved the performance of female chess players across the country (but not outside the country), presumably, as the authors note, because this created an incentive to invest in the skill of female chess players.

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

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

Agan, A. Y., B. Cowgill, and L. K. Gee (2021): "Salary history and employer demand: Evidence from a two-sided audit," Tech. rep., National Bureau of Economic Research.

Aigner, D. J. and G. 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.

Althoff, L. and H. Reichardt (2022): "Jim Crow and Black economic progress after slavery," *Manuscript, Princeton University*.

Altonji, J. G. and R. M. Blank (1999): "Race and gender in the labor market,"

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

Handbook of labor economics, 3, 3143–3259.

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

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

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

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

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

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

Baron, E. J., J. J. Doyle Jr, N. Emanuel, P. Hull, and J. P. Ryan (2023): "Racial Discrimination in Child Protection," Tech. rep., National Bureau of Economic Research.

Baron, J. and J. C. Hershey (1988): "Outcome bias in decision evaluation." *Journal of personality and social psychology*, 54, 569.

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

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

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.

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.

Bierer, B. E., L. G. Meloney, H. R. Ahmed, and S. A. White (2022): "Advancing the inclusion of underrepresented women in clinical research," *Cell Reports Medicine*, 3.

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

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

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

Bohren, J. A., 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.

Buchmann, N., C. Meyer, and C. D. Sullivan (2023): "Paternalistic Discrimination," Working Paper.

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

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., S. Kostyshak, and P. Saygin (2023): "Choosing and Using Information in Evaluation Decisions," Tech. rep., mimeo.

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

Conway, J., R. Mill, and L. C. Stein (2023): "The Gendered Impacts of Perceived Skin Tone: Evidence from African American Siblings in 1870–1940," .

Cook, C., R. Diamond, J. V. Hall, J. A. List, and P. Oyer (2021): "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers," *The Review of Economic Studies*, 88, 2210–2238.

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

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

Crawley, L. M., D. K. Ahn, and M. A. Winkleby (2008): "Perceived medical discrimination and cancer screening behaviors of racial and ethnic minority adults," *Cancer Epidemiology Biomarkers & Prevention*, 17, 1937–1944.

Danz, D., L. Vesterlund, and A. J. Wilson (2022): "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*.

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.

De Sousa, J. and M. Niederle (2022): "Trickle-down effects of affirmative action: A case study in France," Tech. rep., National Bureau of Economic Research.

Derenoncourt, E., C. H. Kim, M. Kuhn, and M. Schularick (2022): "Wealth of two nations: The US racial wealth gap, 1860-2020," Tech. rep., National Bureau of Economic Research.

Dianat, A., F. Echenique, and L. Yariv (2022): "Statistical discrimination and affirmative action in the lab," *Games and Economic Behavior*, 132, 41–58.

Eli, S. J., T. D. Logan, and B. Miloucheva (2023): "The Enduring Effects of Racial Discrimination on Income and Health," *Journal of Economic Literature*, 61, 924–940.

Esponda, I., R. Oprea, and S. Yuksel (2023): "Seeing what is representative," *The Quarterly Journal of Economics*, qjad020.

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.

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.*

Gawai, V. P. and J. D. Foltz (2023): "Discrimination in Science: Salaries of Foreign and US Born Land-Grant University Scientists," .

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

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.*

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

Jerant, A. F., J. J. Fenton, and P. Franks (2008): "Determinants of racial/ethnic colorectal cancer screening disparities," *Archives of internal medicine*, 168, 1317–1324.

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

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

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

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

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

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

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

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

Kupfer, S. S., S. McCaffrey, and K. E. Kim (2006): "Racial and gender disparities in hereditary colorectal cancer risk assessment: the role of family history." *Journal of Cancer Education*, 21.

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

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

Lodermeier, A. (2023): "Racial Discrimination in Eviction Filing," .

Madarász, K. (2012): "Information projection: Model and applications," *The Review of Economic Studies*, 79, 961–985.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

Pinkston, J. C. (2003): "Screening discrimination and the determinants of wages," *Labour Economics*, 10, 643–658.

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*.

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.

Small, M. L. and D. Pager (2020): "Sociological perspectives on racial discrimination," *Journal of Economic Perspectives*, 34, 49–67.

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

Williams, J. A., T. D. Logan, and B. L. Hardy (2021): "The persistence of historical racial violence and political suppression: Implications for contemporary regional inequality," *The ANNALS of the American Academy of Political and Social Science*, 694, 92–107.

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.*

Zink, A., Z. Obermeyer, and E. Pierson (2023): "Race Corrections in Clinical Models: Examining Family History and Cancer Risk," *medRxiv*, 2023–03.

Zivin, J. S. G. and G. Singer (2023): "Disparities in pollution capitalization rates: The role of direct and systemic discrimination," Tech. rep., National Bureau of Economic Research.

A Related Literature

Our framework 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 (Small and Pager 2020), 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.[50] 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 (e.g., Aaronson, Hartley, and Mazumder (2021)). 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.[51]

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—

[50] 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.

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

as in *Griggs v. Duke Power Co. (1970)*.[52] A facially neutral practice may therefore be found to be discriminatory under this doctrine even in the absence of explicit categorization or animus. This notion of discrimination contrasts with the disparate treatment doctrine, which prohibits policies or practices motivated by a discriminatory purpose. Typically, proof of discriminatory intent is required for the finding of disparate treatment.[53]

A systemic perspective is also 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 et al. 2018; Kasy and Abebe 2021; Gebru 2020; Buolamwini 2022; Arnold, Dobbie, and Hull 2021). An algorithm which does not directly use protected characteristics may nevertheless return systematically disparate outcome predictions or treatment recommendations among equally qualified individuals. The literature studies how interlocking systems of data collection, model fitting, and human-algorithm decision-making may generate such disparities.

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

Our work also adds to the long literature on direct discrimination in economics, which is typically modeled as a causal effect of group membership on treatment.[54] Theoretical sources of direct discrimination include individual preferences or beliefs. In the canonical framework of taste-based discrimination, differential treatment emerges because individuals derive disutility from interacting with or providing services to members of a particular group (Becker 1957). In models of belief-based discrimination, differential treatment emerges because a decision-relevant statistic (such as labor market productivity) is unobserved, and there are group-based differences in beliefs about its distribution (Phelps 1972; Arrow 1973; Aigner and Cain 1977). While belief differences have traditionally been assumed to stem from true differences in the distributions, a recent literature has considered the role of inaccurate beliefs in driving direct discrimination (Bohren et al. 2022; Barron, Ditlmann, Gehrig, and Schweighofer-Kodritsch 2020; Hübert and Little 2020). These differences may stem from a lack of information or biased stereotypes (Bordalo, Coffman, Gennaioli, and Shleifer 2016; Coffman, Exley, and Niederle 2021; Bordalo et al. 2019; Fiske 1998), which again lead to

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

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

[54] Notable exceptions to the typical focus on direct discrimination in economics include Neal and Johnson (1996), Glover, Pallais, and Pariente (2017), List (2004), Cook (2014), Hurst et al. (2021), and Sarsons (2019). In Section 3.4 we discuss how the model of Coate and Loury (1993) captures a specific source of systemic discrimination in our framework.

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, Imas, and Rosenberg 2019) and the heterogeneity in discrimination across institutions (Kline, Rose, and Walters 2021).[55] A parallel empirical literature has developed tools to distinguish different economic theories of discrimination. Recent advances involve outcome tests of racial bias, in both observational (Knowles, Persico, and Todd 2001; Grau and Vergara 2021) and quasi-experimental data (Arnold, Dobbie, and Yang 2018; Hull 2021).

As also noted in Small and Pager (2020), the systemic perspective suggests that standard tools for measuring direct discrimination miss an important component. Efforts to model and measure causation at any particular juncture and within a specific domain can substantially understate the cumulative impact of discrimination across domains or time. We contribute to the economics literature by expanding the tools for studying such forms of discrimination. Additionally, our framework offers new interpretations for previously documented group-based disparities. For example, evidence for a reversal of direct discrimination over time—such as the ones documented in Bohren et al. (2019) and Mengel, Sauermann, and Zölitz (2019)—may not imply that total discrimination has been mitigated or reversed. If, as argued, biased evaluators drive initial discrimination in the pipeline, the group that ends up being favored may still face substantial total discrimination when conditioning on underlying qualifications.[56]

A small but growing literature in economics has examined the impact of previous direct discrimination on subsequent disparities. Cook (2014) and Williams, Logan, and Hardy (2021) study the long-run effects of racial violence on innovation and regional inequality,

[55]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.

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

respectively. Eli, Logan, and Miloucheva (2023) and Derenoncourt, Kim, Kuhn, and Schularick (2022) review and examine the impact of historical discriminatory practices on the evolution of the racial wealth gap.

A series of papers have built directly on our definitions and framework to measure and classify direct, systemic, and total discrimination. Althoff and Reichardt (2022) measure the systemic components of disparities that stem from racially oppressive institutions—slavery and Jim Crow laws. Baron et al. (2023) examine discrimination in foster care through the investigator-screener relationship, finding that systemic discrimination generated by screeners accounts for a substantial proportion of the resulting total discrimination. Zivin and Singer (2023) study racial differences in home values as a function of pollution exposure, concluding that 75% of the disparity was driven by systemic discrimination in complementary amenities. Lodermeier (2023) applies our framework to the study of eviction rates, finding that the substantial racial disparity is likely caused by direct rather than systemic discrimination. Gawai and Foltz (2023) look at the impact of country of birth on income in academia and find significant total discrimination. They identify two-thirds of that disparity to be driven by systemic discrimination. Finally, Buchmann et al. (2023) study a form of anticipated systemic discrimination where employers are less likely to hire women due to gender-based disparities in safety outside of the job, which they term *paternalistic discrimination*. They find that eliminating this type of discrimination would reduce the gender employment gap by 24% and increase female wages by 21% in their setting.

B Additional Applications

B.1 Signal Inflation with Non-Exogenous Groups

This experiment illustrates how systemic discrimination can arise from signal inflation in a setting with non-exogenously assigned groups (Male or Female). In the experiment, a pool of workers faced evaluations from two sets of managers, Recruiters and Hiring Managers. Recruiters generated initial evaluations of Workers based on a productivity signal and their self-identified gender; Hiring Managers evaluated workers based on group identity and the evaluation of the recruiter. Worker qualification was chosen so that there is no systemic gender discrimination in Recruiter evaluations: total discrimination equals direct discrimination in this stage. Direct discrimination by Recruiters could lead to (informational) systemic discrimination in Hiring Manager' evaluations, alongside additional direct discrimination by Hiring Managers.

B.1.1 Experimental Setup

Workers: 100 participants were randomly assigned to the role of Worker. Each Worker completed two sets of tasks (A and B) and provided basic demographic information including self-reported group identity G_i (either male m or female f). Each task consisted of a test of the Worker's basic math, business, and history knowledge, with 10 randomly selected questions from these subjects. A Worker's performance on each task was defined as the number of questions she answered correctly. We restrict attention to Workers with a task-A performance in $\mathcal{S}^R = \{2, 3, 4, 5, 6\}$ in order to ensure enough data for each gender.

Recruiters: 201 participants were randomly assigned to the role of Recruiter and given a budget of 10 experimental units (10 EU=\$1 USD). Each Recruiter was shown information about two Workers and reported their highest willingness to pay to hire each. Specifically, Recruiters were shown the task-A performance of the Worker, which constituted their signal S_i^R, as well as the Worker's gender G_i. After viewing S_i^R and G_i, Recruiters were asked to state their willingness to pay to hire Worker i in the range of 0-10 EUs. This willingness to pay constituted the Recruiter action, A_i^R, with $\mathcal{A} = \{0, \ldots, 10\}$. Recruiter wage offers were then accepted or rejected according to the Becker-DeGroot-Marschak mechanism to incentivize truthful reporting: if a Worker was hired, Recruiters received 1 EU for each question the Worker answered correctly on task B minus the wage. If the Worker was not hired, the Recruiter did not pay anything and kept their endowment.[57]

Hiring Managers: 504 participants were randomly assigned to the role of Hiring Manager and also given a budget of 10 EU. Each Hiring Manager was shown the gender and Recruiter wage offer of a Worker.[58] Formally, each Hiring Manager observed signal $S_i^H \equiv A_{ik}^R$ for some Recruiter k assigned to Worker i, with $\mathcal{S}^H = \{0, \ldots, 10\}$. Hiring Managers then stated their maximum willingness to pay to hire the Worker using the same Becker-DeGroot-Marschak mechanism as the Recruiters. We denote the Hiring Manager's action (wage offer) as A_i^H, with $\mathcal{A} = \{0, \ldots, 10\}$ as before.

B.1.2 Results

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

Workers: There were no significant gender differences in Worker performance on either task. On average, Workers completed 3.57 questions correctly on task A and 3.53 questions correctly on task B. Regressing overall performance (the sum of performance on both tasks)

[57]Here and in Appendix B.2 we censor earnings at zero so that they could not be negative. Both Recruiters and Hiring Managers saw examples of the mechanism, examples of the task faced by the Workers, and passed comprehension checks before making wage offers.

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

Table 4. Signal Inflation: Recruiter and Hiring Manager Wage Offers

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

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

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

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

Recruiters directly discriminated against female Workers. The average offered wage was 5.23. Column 1 of Table 4 shows that male Workers were on average offered a 0.49 higher wage than female Workers ($p < 0.01$).[59] 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 ($p < 0.01$).[60] While this data alone cannot be used to disentangle preference and belief-based sources of direct discrimination, it is consistent with work showing inaccurate beliefs or stereotypes as drivers of discrimination in similar settings (Bordalo et al. 2019; Bohren et al. 2019).

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

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

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

Hiring Managers discriminated against female Workers. The average Hiring Manager wage offer was 5.50. Column 3 of Table 4 shows that male Workers were on average offered a 0.94 higher wage than female Workers ($p < 0.01$). This disparity captures corresponds to roughly 0.39 standard deviations of Hiring Manager wage offers.

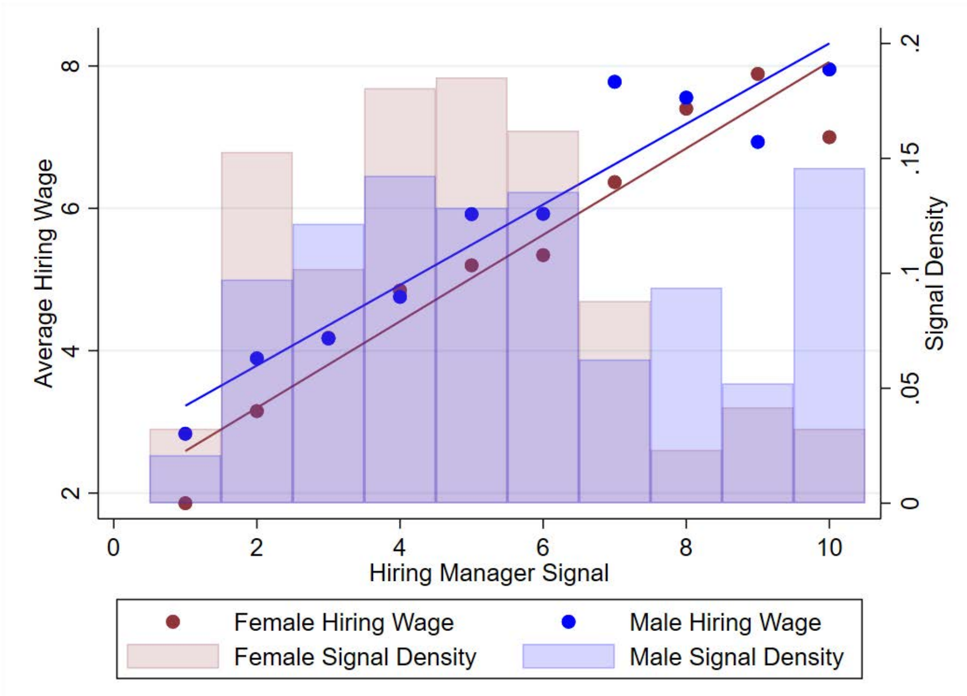
Column 4 of Table 4 further suggests that much of the discrimination by Hiring Managers is systemic. Controlling for the Hiring Manager signal (i.e., the Recruiter wage offer) decreases the gender coefficient to 0.41 ($p = 0.02$). Interestingly, both the gender coefficient and productivity signal coefficient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

Figure 5 illustrates the two sources of Hiring Manager discrimination. The scatter plot shows average Hiring Manager wages as a function of the Worker's gender and productivity signal. The lines of best fit show a positive relationship between the signal and wage for both genders. This relationship is shifted upward for male Workers, illustrating direct discrimination: conditional on seeing the same signal, a male Worker received a higher wage than a female Worker. Importantly, however, the *distribution* of productivity signals differs by gender: male Workers tend to have higher signals than female Workers, due to direct discrimination in initial evaluations. Since higher signals lead to higher wages from the Hiring Managers (the upward sloping lines), this pattern leads to systemic discrimination.

We now quantify systemic discrimination using the decompositions in Section 4.2. We first estimate total Hiring Manager discrimination $\Delta(y^0)$ by comparing male and female wage offers for each task-A performance level. We then estimate Hiring Managers' average direct discrimination with a given task-A performance, $E[\tau_i \mid G_i = w, Y_i^0 = y^0]$, by averaging gender disparities across each Hiring Manager signal realization according to the distribution each task-A performance induces over the Hiring Manager signal (i.e., the Recruiter wage offer). Per Equation (7), subtracting this estimate of direct discrimination from the estimate of total discrimination yields an estimate of systemic discrimination at each qualification level y. We then average these measures of total, direct, and systemic discrimination over the marginal distribution of Worker qualification by gender before equal-weighting these gender-specific averages.[61] We similarly decompose total discrimination into the alternative measures of direct and systemic components in Equations (8) and (9).

Table 5 confirms significant systemic discrimination in Hiring Manager wage offers. Estimated total discrimination against female Workers averages to 0.90, similar to the regression

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



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

FIGURE 5. Signal Inflation: Hiring Manager Wage Offers by Worker Gender and Signal

estimate in Column 3 of Table 4. Estimated average direct discrimination is around 0.41 for each decomposition, similar to the regression estimate in Column 4 of Table 4. Estimated systemic discrimination is around 0.49 for each of the three decompositions. The majority (54%) of discrimination thus comes from signal inflation.

Our lab experiment illustrates 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: direct Hiring Manager discrimination, which conditions on the non-gender signal, was much smaller than total discrimination. The study also illustrates 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. Therefore policies which aim to eliminate direct discrimination through contact (Rao 2019; Paluck, Green, and Green 2019) or correcting beliefs (Bohren et al. 2022) may still allow discrimination to persist through systemic factors.

TABLE 5. Signal Inflation: Discrimination Decomposition

	(1)	(2)	(3)
Total	0.90***	0.90***	0.90***
	(0.22)	(0.22)	(0.22)
Average Direct	0.41*	0.41**	0.41**
	(0.23)	(0.18)	(0.19)
Systemic	0.49**	0.49***	0.49***
	(0.22)	(0.16)	(0.18)
# Observations	504	504	504
Decomposition Method	Eq. (7)	Eq. (8)	Eq. (9)

Notes: This table reports estimates of each measure of discrimination in Equations (7) to (9) for Hiring Manager wage offers, averaged by an equal-weighted distribution of task-A scores for male and female Workers (the qualification). Total discrimination is measured by the average difference in wages for male versus female Workers with a given task-A score. The sample includes 504 Hiring Managers, each evaluating one Worker. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

B.2 Screening

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

Both the theoretical example and experiment show that canonical statistical discrimination models may not capture the full extent of (total) discrimination stemming from differences in the signaling technology. They also show 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. Finally, we show 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.

B.2.1 Theoretical Example

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

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

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

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

where $\Phi(\cdot)$ gives the standard normal distribution.[62] Since $\eta_b \neq \eta_w$, this expression is non-zero unless $y = t\frac{1+\eta_g}{\eta_g}$. Therefore, there is systemic discrimination almost everywhere in the productivity distribution, stemming from the differential probabilities of the signal being above a given cutoff for equally productive group-w versus group-b workers.

Systemic discrimination in this screening action is heterogeneous across worker productivity levels. With $\eta_w > \eta_b > 0$, the systemic discrimination hurts group-b workers at high levels of productivity (where $y > t\frac{1+\eta_g}{\eta_g}$) and *favors* group-$b$ workers at low levels of pro-

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

ductivity (where $y < t\frac{1+\eta_g}{\eta_g}$) since $\Phi(\cdot)$ is strictly increasing. Intuitively, having a higher signal variance makes low-productivity group-b workers more likely to have a signal above the effective threshold by chance, while high-productivity group-b workers are more likely to generate a signal below the threshold by chance.

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

This theoretical example highlights how examining screening discrimination with only a direct measure of discrimination may miss an important component of how differential signal precision impacts total discrimination.

B.2.2 Experiment

We proceed to illustrate screening discrimination empirically in an online labor market, using a setup similar to the one in Appendix B.1.

Experimental Setup

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

A new group of 501 Hiring Managers saw one Worker's profile after their evaluation by a Recruiter, along with the Worker's gender. They were shown information on the Worker's task-A performance only if the Recruiter had chosen to hire them; otherwise they saw no performance information. Therefore, $\mathcal{S}^H = \{\emptyset, 2, 3, 4, 5, 6\}$. Hiring Managers then made a binary decision of whether or not to hire the Worker. If the Worker was hired, the Hiring Manager received a bonus corresponding to their task-B performance; otherwise, the Hiring Manager received 4 dollars with certainty.

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

Results

As before, we measure systemic and total discrimination with respect to task-A performance, $Y_i^0 = S_i^R$, with $\mathcal{Y}^0 = \{2, 3, 4, 5, 6\}$. Since this qualification measure coincides with the Recruiter signal, any discrimination in the Recruiter stage is direct. Discrimination in the Hiring Manager stage can again be direct or systemic. Per Appendix B.2.1, we expect the differences in signal informativeness to lead to heterogeneity in systemic discrimination by qualification.

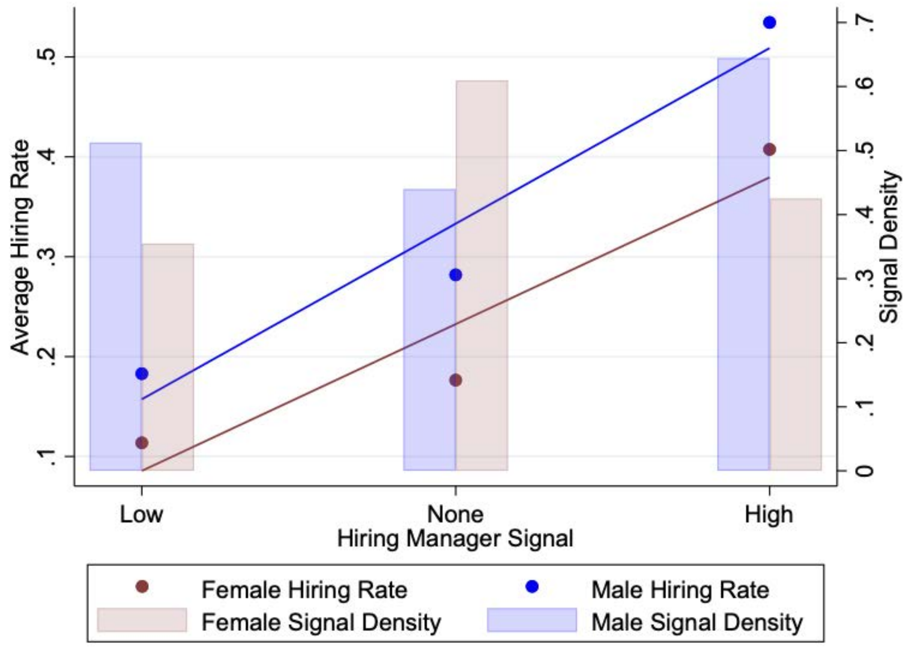
Recruiters: Recruiters directly discriminated against female Workers. The hiring rate for male Workers was 28 percentage points higher than for female Workers ($p < 0.01$), who were hired at a rate of 36%.[63] Given the lack of gender-based performance differences, as reported in Appendix B.1.2, this disparity in hiring rates is not consistent with accurate statistical discrimination. Therefore, Recruiter direct discrimination again stems from either biased preferences or beliefs.

Hiring Managers: Hiring Managers discriminated against female Workers. On average, male Workers were hired at a 9 percentage point higher rate than female Workers ($p = 0.02$), who were hired at a rate of 0.22. However, this average effect masks important heterogeneity. Among Workers with low (below-median) qualification levels, male Workers were hired at an insignificant 4 percentage point higher rate ($p = 0.43$).[64] Among Workers with high (above-median) qualification levels, male Workers were hired at a significant 23 percentage point higher rate ($p < 0.01$).

Figure 6 illustrates the reason for this heterogeneity in total discrimination. Similar to Figure 5, the scatter plot shows the average Hiring Manager actions conditional on the signal (or lack thereof) and the Worker's gender. As before, the lines of best fit show a positive relationship between the signal and the probability of getting hired for both groups: Hiring Managers were more likely to hire a Worker after seeing a high signal than a low signal, with the hiring rate for no signal laying in between. Conditional hiring rates are shifted upward for male Workers, illustrating direct discrimination. Importantly, however, the distribution of signals seen by Managers also differs by gender: direct discrimination by Recruiters made Managers more likely to see both low and high signals from male Workers

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

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



Notes: This figure plots average Hiring Manager hiring rates (left y-axis) and signal probabilities (right y-axis) by productivity signal for female and male workers, where high versus low signal corresponded to either above and equal to or below the median (3), respectively. Gender differences in the hiring rates for a given signal illustrates direct discrimination, while gender differences in the signal probability illustrates the source of systemic discrimination.

FIGURE 6. Screening: Hiring Manager Hiring Rate and Signals

than female Workers, with female Workers being much more likely to have an uninformative signal. Given the upward-sloping lines, female Workers with high qualification levels were likely to be hurt by systemic discrimination, while female Workers with low qualification levels were likely to be helped by it.

We quantify total, direct, and systemic discrimination in Hiring Manager actions using the decompositions in Section 4.2. We estimate Hiring Manager total discrimination $\Delta(y^0)$ by comparing male and female hiring rates based on task-A performance. We then estimate the Hiring Manager average direct discrimination $\overline{\tau}(w, y^0)$ faced by male Workers with a given task-A performance by averaging gender disparities across each Hiring Manager signal realization according to the distribution each task-A performance induces over this signal. Subtracting this estimate of from the estimate of total discrimination yields an estimate of the measure of systemic discrimination.[65] We average these measures over the distribution of task-A performance as before, separately for Workers with low (below-median) and high (above-median) qualification levels.

Table 6 confirms the heterogeneity in systemic discrimination faced by women with different qualification levels. For highly qualified women, total discrimination is estimated as

[65]Here we use the "average" decomposition, Equation (8). The other decompositions give similar results.

Table 6. Screening: Discrimination Decomposition

	High Qualification (1)	Low Qualification (2)	Difference (3)
Total	0.24***	0.03	0.21***
	(0.06)	(0.04)	(0.07)
Average Direct	0.15***	0.07**	0.08
	(0.05)	(0.04)	(0.05)
Systemic	0.09**	-0.04	0.13**
	(0.04)	(0.03)	(0.06)
# Observations	501	501	501

Notes: This table reports estimates of each measure of discrimination in Equation (9) for Hiring Manager hiring rates, averaged by an equal-weighted distribution of task-A scores for male and female Workers in the given qualification bin, where High corresponds to above or equal to the median (3) and Low corresponds to below the median. Total discrimination is measured by the average difference in hiring rates among male versus female Workers with a given task-A score. The sample includes 501 Hiring Managers, each evaluating one Worker. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

a significant 0.24. Our decomposition shows this is driven by a combination of significant direct (0.15) and systemic discrimination (0.09). In contrast, total discrimination among workers with a low qualification is small and insignificant (0.03), despite significant direct discrimination. The reason is a small degree of negative systemic discrimination among less qualified Workers (-0.04). Consistent with the model in Appendix B.2.1, the gap in systemic discrimination across qualification levels is significant ($p = 0.04$).

C Additional Analyses from Section 5

C.1 Additional Analysis from Section 5.2.

We first present the two other versions of the decomposition with productivity as the qualification. Note that all three decompositions are equivalent when the signal is the qualification, as the average direct term is not averaged across multiple signal realizations and the systemic term is zero. We then present the regerssion results used to estimate total discrimination with the signal as the qualification. Finally, we present an analysis of prediction bias.

Other Decompositions. Table 7 presents the measures of discrimination from Eq. (7) and Eq. (8) for Recommenders, while Table 8 does so for Evaluators.

Table 7. Mitigation Study: Additional Decompositions for Recommenders

	Group-Blind		Informed	
	(1)	(2)	(3)	(4)
Qualification: Productivity				
Total	-0.85***	-0.85***	-0.46***	-0.46***
	(0.09)	(0.09)	(0.12)	(0.12)
Average Direct	-0.05	-0.03	0.07**	0.08**
	(0.08)	(0.07)	(0.03)	(0.03)
Systemic	-0.80***	-0.82***	-0.53***	-0.53***
	(0.10)	(0.10)	(0.13)	(0.13)
# Observations	780	780	720	720
Decomposition Method	Eq. (7)	Eq. (8)	Eq. (7)	Eq. (8)

Notes: This table reports estimates of each measure of discrimination in either Eq. (7) or Eq. (8), averaged by the population distribution of productivity. The sample includes 300 Recommenders, each evaluating five Workers. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 8. Mitigation Study: Additional Decompositions for Evaluators

	Group-Blind		Informed	
	(1)	(2)	(3)	(4)
Qualification: Productivity				
Total	-0.95***	-0.95***	-1.04***	-1.04***
	(0.08)	(0.08)	(0.11)	(0.11)
Avgerage Direct	-0.04	0.02	0.10	0.13
	(0.08)	(0.07)	(0.15)	(0.13)
Systemic	-0.91***	-0.96***	-1.14***	-1.17***
	(0.08)	(0.08)	(0.12)	(0.09)
# Observations	1055	1055	945	945
Decomposition Method	Eq. (7)	Eq. (8)	Eq. (7)	Eq. (8)

Notes: This table reports estimates of each measure of discrimination in either Eq. (7) or Eq. (8), averaged by the population distribution of productivity. The sample includes 400 Evaluators, each evaluating five Workers. Robust standard errors, obtained from a weighted bootstrap, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Regression to Estimate Direct Discrimination. Table 9 presents results from a linear regression of the evaluation on group, the signal, and their interaction. In the case of Group-Blind Recommenders, the coefficients are small and insignificant on Group b—which corresponds to direct discrimination at signal $S_i = 0$, i.e., $\tau(0)$—and the interaction

Table 9. Mitigation Study: Regression Analysis of Evaluations

	Recommender		Evaluator	
	Group-Blind (1)	Informed (2)	Group-Blind (3)	Informed (4)
Group b	-0.064	0.951^{***}	-0.339	-0.110
	(0.188)	(0.193)	(0.238)	(0.260)
Signal S	0.466^{***}	0.625^{***}	0.613^{***}	0.720^{***}
	(0.031)	(0.029)	(0.049)	(0.047)
Group b × Signal S	0.018	-0.073^{*}	0.081	0.049
	(0.036)	(0.038)	(0.053)	(0.054)
Constant	2.081^{***}	0.981^{***}	1.800^{***}	1.278^{***}
	(0.168)	(0.165)	(0.242)	(0.239)
# Observations	780	720	1055	945

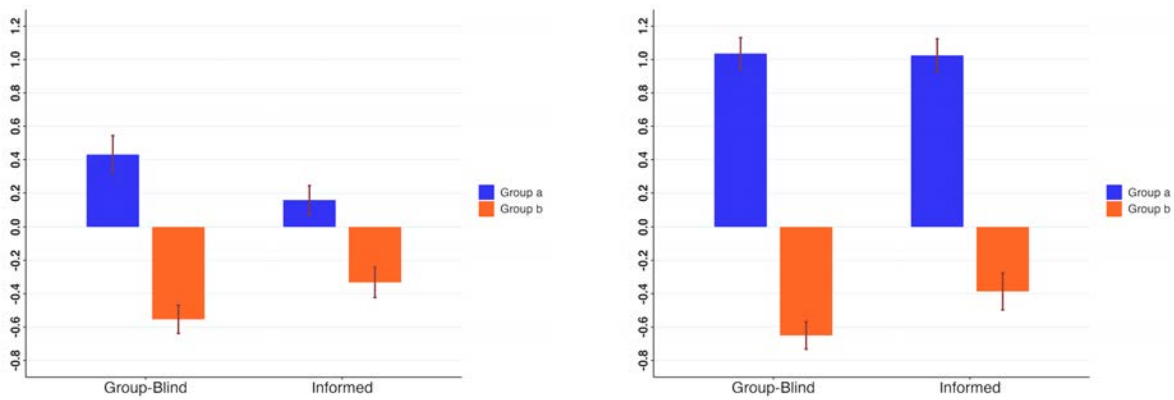
Notes: This table reports coefficients from regressing Recommender or Evaluator evaluations on Worker group, the signal, and their interaction. Standard errors, clustered at the Recommender or Evaluator level, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

term—which corresponds to the slope of direct discrimination with respect to the signal, i.e. $d\tau(s)/ds$ (Column 1 of Table 9).

In the case of Informed Recommenders, the coefficient on Group b is positive and significant, suggesting substantial *reverse* direct discrimination that favored Group b workers with signal $S_i = 0$ by 0.95 points (Column 3 of Table 9). The coefficient on the interaction term is negative and significant, suggesting less reverse discrimination at higher signal realizations. But even at a high signal of $S_i = 10$, direct discrimination favors Group b by 0.22.

Prediction Bias. To assess the direction of evaluation bias, we examine the difference between the evaluation and actual productivity, $A_i - Y_i^*$, which we refer to as the *Prediction Bias* (PB). Fig. 7(a) plots the PB for each group and treatment. Unsurprisingly, Group-Blind Recommenders select evaluations that are significantly upwardly biased for Group a participants and downwardly biased for Group b participants. Informed Recommenders remain biased in the same direction for each group, though less so than in the Group-Blind treatment. Thus, while information somewhat mitigates the impact of signal inflation, the prediction bias remains sizable and significant. Moreover, the magnitude of the bias is larger for Group b than Group a, suggesting that it is more difficult to adjust for negatively slanted information than inflated information. Table 10 presents regression analyses of the bias.

In the case of Group-Blind Evaluators, prediction bias compounded across rounds, particularly in the case of upward bias for Group a. As shown in Fig. 7, Group-Blind Evaluators exhibited more than twice as much positive bias towards Group a as Group-Blind Rec-



(a) Recommenders (b) Evaluators

Note: Whiskers indicate +/- one robust standard error.

FIGURE 7. Mitigation Study: Prediction Bias by Group and Treatment

ommenders. Notably, if Group-Blind Evaluators simply reported the Group-Blind Recommenders' evaluations (the optimal strategy if Evaluators believed that Recommenders were using information optimally), then Group-Blind Evaluators and Group-Blind Recommenders would have the same prediction bias.

As shown in Fig. 7(a), Informed Evaluator evaluations of Group a workers have a similar magnitude positive bias as the Group-Blind treatment, while evaluations of Group b have a somewhat smaller negative bias. In contrast, Recommender prediction bias decreased for both groups between Group-Blind and Informed treatments. Table 10 presents these results in regression form, showing that while information about an exogenous signal-generating process significantly decreased the extent of prediction bias (Recommenders), information about an endogenous signal-generating process did not (Evaluators).

C.2 Analysis of Group-Blind Treatment from Section 5.3.

In this treatment, participants chose how many tables to solve before learning group assignment. They were told that group would be randomly assigned with equal probability of a and b. Participants in this treatment chose to complete 5.39 tables on average, with a standard error of 0.33. In terms of (hypothetical) payment, Company 1 paid them \$0.57 on average, with a standard error of 0.05. Company 2 paid them \$0.51 on average, with a standard error of 0.04. Thus, these productivity choices and payments lie between the average productivity choices and payments when participants first observe their group.

Table 10. Mitigation Study: Prediction Bias Regressions

	Recommenders (1)	Evaluators (2)
Constant	0.430***	1.036***
	(0.114)	(0.095)
Group b	−0.984***	−1.686***
	(0.124)	(0.114)
Informed	−0.273*	−0.011
	(0.142)	(0.137)
Group b × Informed	0.494***	0.275
	(0.174)	(0.185)
# Observations	1,500	2,000

Notes: This table reports coefficients from regressing Recommender and Evaluator prediction bias on Worker group, Recommender or Evaluator treatment, and their interaction. Column 1 includes 300 Recommenders, each evaluating five Workers. Column 2 include 400 Evaluators, each evaluating five Workers. Standard errors, clustered at the Recommender or Evaluator level, are reported in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.