Inaccurate Statistical Discrimination: An Identification Problem

J. Aislinn Bohren, Kareem Haggag, Alex Imas, and Devin G. Pope*

December 16, 2022

Abstract

This paper studies inaccurate beliefs as a source of discrimination. Economists have typically characterized discrimination as stemming from tastes (preference-based) or accurate statistical (belief-based) sources—a valuable distinction for policy design and welfare analysis. However, in many situations individuals may have inaccurate beliefs about how relevant characteristics—e.g. productivity, signals—are correlated with group identity. A review of the empirical discrimination literature in economics reveals that a small minority of papers—fewer than 7%—consider the possibility of such *inaccurate statistical discrimination.* Using a theoretical framework and an experiment in a labor market setting, we show that not accounting for inaccurate beliefs will lead to a misclassification of discrimination's source. We then outline three methodologies that either fully or partially identify the three potential sources: varying the amount of information presented to evaluators, eliciting their beliefs, and presenting them with accurate information about the relevant distributions. Importantly, the third method can be used to differentiate whether inaccurate beliefs are due to a lack of information or motivated factors.

KEYWORDS: Discrimination, Inaccurate beliefs, Model misspecification

*Bohren: University of Pennsylvania, abohren@gmail.com, Haggag: Anderson School of Business, UCLA, kareem.haggag@anderson.ucla.edu. Imas: Booth School of Business, University of Chicago, alex.imas@chicagobooth.edu. Pope: Booth School of Business, University of Chicago, devin.pope@chicagobooth.edu. We thank Steven Durlauf, Hanming Fang, Alex Frankel, Emir Kamenica, Emily Nix, and seminar participants at Harvard Business School, Harvard Kennedy School, the SaMMF Discrimination in Labor Markets Workshop, Stanford University, UCLA, University of Chicago, University of Melbourne, University of Pennsylvania, University of Southern California, University of Sydney, University of Virginia, Cambridge University and the Virtual Market Design Seminar for helpful comments and suggestions. Cuimin Ba, Byunghoon Kim and Jihong Song provided excellent research assistance. Bohren gratefully acknowledges financial support from NSF grant SES-1851629. The experiment received IRB approval at CMU.

1 Introduction

Discrimination based on group identity has been shown to be prevalent in many important settings, including labor markets, housing markets, credit markets, and online consumer markets (see Bertrand and Duflo (2017) and Charles and Guryan (2011) for reviews). Economists studying direct discrimination—i.e. the causal link between group identity and differential treatment—often also seek to identify its source.[1] Sources are typically categorized into one of two forms. In the case of taste-based discrimination (Becker, 1957), an individual has animus towards members of a particular group and discriminates against them because he receives disutility from providing services to or interacting with members of this group. In the case of accurate statistical discrimination (Phelps, 1972; Aigner and Cain, 1977), differential treatment occurs because productivity is unobserved and a particular group's distribution of productivity is *correctly* perceived to have either a lower mean, differential variance, or differential precision of signal about it, relative to an alternative group.[2]

Although statistical discrimination is typically assumed to be driven by rational expectations, a large literature in psychology and economics has shown that people's beliefs are often incorrect.[3] This motivates the topic of the current paper, which studies the role of inaccurate beliefs about productivity and signal distributions in driving discrimination. Using a theoretical framework and an experiment in a labor market setting, we demonstrate the importance of accounting for inaccurate beliefs when classifying the source of observed disparities. We show that such beliefs often give rise to similar patterns in the data as taste-based sources; in turn, commonly-used methods cannot disentangle inaccurate beliefs and preferences as drivers of discrimination. Moreover, failure to account for inaccurate beliefs can lead to a potential misclassification of discrimination's source. We outline three alternative identification methods: eliciting beliefs, varying the number of signal draws, and providing direct information about the productivity distribution.

A systematic review of the discrimination literature reveals that while a large plurality of papers (61.9%) attempts to differentiate between taste-based versus statistical sources, only a small proportion (10.5%) discusses the possibility of inaccurate beliefs.

[1] Bohren, Hull, and Imas (2022) considers the role of both direct and indirect, i.e. *systemic* sources of discrimination in generating group-based disparities. The current paper focuses on the former channel.

[2] Differences in the productivity distribution may be due to exogenous differences (Phelps, 1972) or part of a self-fulfilling equilibrium (Arrow, 1973).

[3] See for example Kravitz and Platania (1993) and Fiske (1998).

Yet identifying the source of discrimination is important for a myriad of reasons: designing an effective policy intervention to reduce discrimination crucially depends on what drives it,[4] welfare and efficiency analyses differ as a function of the source, and the extent to which competitive markets will eliminate discrimination depends on whether it stems from preferences or beliefs (see Fang and Moro (2011) for review). Additionally, when discrimination stems from inaccurate beliefs or preferences, it can lead to further discrimination by *other* people (or algorithms) who learn from the decisions of the discriminators but are unaware of their bias (Bohren et al., 2022).

To formalize how the possibility of inaccurate beliefs impacts identification, we first develop a theoretical framework for modeling inaccurate statistical discrimination. Consider an evaluator who observes the group identity of a worker as well as a signal about her productivity, then decides whether to hire the worker. *Direct discrimination* occurs when two workers who generate identical signals are evaluated differently based on their group identity. This discrimination can either stem from belief-based partiality—where evaluators have group-dependent beliefs about the productivity and/or the signal distributions—or preference-based partiality—where evaluators have group-dependent preferences over hiring workers of a given expected productivity. The former is typically referred to as statistical discrimination, while the latter is referred to as taste-based discrimination, prejudice, or animus. We expand this standard framework by considering both accurate and inaccurate beliefs about the productivity and signal distributions.

We first characterize the set of preferences and beliefs that result in equivalent discrimination—that is, the same pair of hiring rules. It is readily apparent that a continuum of preference and belief profiles can give rise to equivalent discrimination.[5] Therefore, identifying a given pair of hiring rules does not identify the source of discrimination. Further, we establish that it does not even rule out any form of inaccurate beliefs e.g. inaccurate beliefs about the signal precision versus average productivity, etc. Additional data is necessary to isolate the source of discrimination.

It may be that a researcher also has access to the true productivity and signal distri-

[4]For example, if discrimination stems from inaccurate beliefs, an effective policy response could be providing individuals with information about the correct distributions, whereas such a policy would have no effect when discrimination stems from the other two sources. See for example, Jensen (2010) in the case of inaccurate beliefs about the returns to education, or Bursztyn, González, and Yanagizawa-Drott (2020) in the case of inaccurate beliefs about the beliefs of others, i.e. pluralistic ignorance.

[5]Manski (2004) first illustrated that observed choice behavior could be consistent with multiple sets of preferences and subjective beliefs, and hence, identification of preferences from choice data required strong assumptions, such as rational expectations. He proposed that data on expectations could be used to validate or relax the rational expectations assumption.

butions. In such cases, studies have used a method often referred to as an outcomes-based test that compares evaluation decisions to the true distributions in order to identify the source of discrimination.[6] For example, a researcher may compare differences in lending rates between two groups to differences in their loan default rates. When maintaining the assumption that evaluators have accurate beliefs, this method pins down the source of discrimination. However, identification depends critically on this assumption: without it, we show that the only source that can be ruled out is accurate statistical discrimination, i.e. an evaluator with accurate beliefs and no preference partiality.[7] Moreover, erroneously assuming that an evaluator has accurate beliefs leads a researcher to mistakenly attribute the share of discrimination arising from inaccurate beliefs to preference partiality. Depending on whether an evaluator's inaccurate beliefs increase or decrease discrimination relative to an evaluator with accurate beliefs and the same preferences, the researcher will over- or underestimate the degree to which the evaluator has preferences that favor one of the groups.

As an example of this misclassification, consider a study that finds evidence for discrimination and measures productivity outcomes. If the researcher observes that both groups have identical productivity and signal distributions and assumes that evaluators have correct beliefs about these distributions, she concludes that the source of the observed discrimination must be preference-based. However, an alternative explanation is that evaluators have incorrect beliefs, which lead to inaccurate statistical discrimination. Without further data, it is impossible to distinguish between these two explanations.

We then outline alternative methods for identifying the source of discrimination in this expanded framework. One method, as proposed by Manski (2004), is to directly collect data on the subjective beliefs of evaluators. Combined with observing the evaluation decisions and signals, this identifies preferences. Data on the true distributions is also required in order to determine whether beliefs are accurate.[8] In many settings, eliciting

[6]This commonly-used method has been employed in many domains, including lending, policing, and bail decisions (Pope and Sydnor, 2011; Knowles, Persico, and Todd, 2001; Antonovics and Knight, 2009; Arnold, Dobbie, and Hull, 2022).

[7]Recent work by Arnold, Dobbie, and Yang (2018) and Grau and Vergara (2021) consider a different type of outcome-based test that does not assume that the researcher observes the decision-maker's signals. Using IV and marginal treatment effect (MTE) methods, these tests can also reject accurate statistical discrimination (Hull, 2021), but the identification problem outlined in the current paper remains.

[8]A growing empirical literature elicits expectations to separate preferences from subjective expectations, including birth control choices (Delavande, 2008), college major choices (Wiswall and Zafar, 2015), and secondary education choices (Giustinelli, 2016).

beliefs will not be feasible. An alternative method is to manipulate the signal precision by varying the number of signal draws observed by evaluators. For example, one could vary the number of recommendation letters for a job candidate or the number of reviews on a platform like TaskRabbit. We demonstrate that this method can partially identify the source of discrimination: it identifies the extent of preference-based partiality, but cannot distinguish between different forms of belief-based partiality (i.e. different beliefs about average productivity versus signal precision). Importantly, this method requires multiple signals from the same domain (e.g. reviews from the same population of evaluators); if the signals are from different domains (e.g. SAT scores and education history), then the identification problem persists.[9]

We next demonstrate the identification issue and alternative methods in a stylized hiring experiment. Participants are recruited and assigned to the role of either "worker" or "employer". Workers created profiles that included a variety of characteristics, such as their country of origin (US vs. India), gender, and age, along with other information such as their beverage and movie preferences. They then completed a series of logical reasoning questions. Employers were shown the profiles of 20 workers and asked the maximum wage they would be willing to pay to hire each worker. The employer's payoff depended on her offered wage and how many questions the worker answered correctly.

We find that employers discriminate based on the worker's country of origin and gender: Americans and females received systematically lower wage offers than Indians and males. According to the standard classification, the observed discrimination is generated by two potential sources. Employers may offer lower wages to American and female workers because they believed that members of those groups answer fewer questions correctly on average than Indian and male workers. Since they lack information on the productivity of any given worker, employers used these group statistics to inform their compensation decisions. Alternatively, employers may be prejudiced towards members of the discriminated group and offered them lower wages because they did not want to reward them.

[9] Another approach which we do not study in this paper derives predictions from a specific structural model of biased beliefs and takes these predictions to the data. For example, Arnold et al. (2018) compare the distributions of pre-trial misconduct of marginal black and white defendants. They argue that the distributional differences are consistent with bail judges holding incorrect stereotypes about the release risk of black defendants. Bohren, Imas, and Rosenberg (2019) model how discrimination evolves across evaluation rounds in a social learning setting, and argue that the observed dynamics are consistent with discrimination driven by inaccurate belief partiality but inconsistent with accurate belief partiality or preference partiality.

As discussed above, outcomes-based tests are often used to distinguish between these sources by comparing the compensation decisions to the "ground truth"—the true performance distributions by group. Our experiment allows us to measure the "ground truth" by comparing the number of questions answered correctly across the various groups. We find that, if anything, Americans slightly outperform Indians on the task (although the difference is not statistically significant), while females perform less well than males. Under the assumption of accurate beliefs, we would conclude that the discrimination against Americans is due to preference partiality. Further, because the level of discrimination against females is substantially smaller than the actual gap in performance, this approach would conclude that evaluators have preference partiality against men.

However, an alternative explanation is that individuals have no preference partiality towards or against a particular group, but rather, have inaccurate beliefs about the respective performance distributions. To identify this channel, we elicited the beliefs of employers and compared them to the "ground truth." Consistent with inaccurate statistical discrimination, employers mistakenly predicted that American workers perform much worse than their Indian counterparts, and that female workers only slightly underperform relative to males. Accounting for these inaccurate beliefs substantially changes the inferred source of discrimination. What was originally classified as preference-based discrimination *in favor of* Indians is mostly explained by mistaken beliefs—if anything, the preference-based channel goes slightly *against* Indian workers. Similarly, a large portion of the gender gap in wages can be explained by inaccurate statistical discrimination.

The line between inaccurate beliefs and animus may sometimes be blurry. For example, individuals may develop inaccurate beliefs *because* they have animus against members of a particular group. We propose that these channels can be separately identified through the provision of information about the relevant distributions. Specifically, if agents are provided with credible information on how the productivity or signal distributions vary by group, those with inaccurate beliefs should update their beliefs and adjust their behavior accordingly. However, if mistaken beliefs merely mask an underlying animus, then agents are unlikely to change their behavior in response to such information. We implement this method in our experiment by providing employers with information on average performance by gender, nationality, and age. After receiving this information, participants were asked to make wage offers to 10 additional workers. We find that employers significantly changed their wage offers in the direction consistent with correcting their beliefs. This methodology is portable outside of our stylized experimental setting

as a way to identify animus-driven inaccurate beliefs versus inaccurate beliefs stemming from inexperience or a lack of information.

The paper proceeds as follows. Section 2 presents a review of the economics literature on discrimination, demonstrating that few papers consider mistaken beliefs when attempting to isolate its source. Section 3 outlines our theoretical framework and results. Section 4 illustrates these findings in a stylized hiring experiment. Section 5 concludes.

2 Survey of the Literature

We conducted a systematic survey of the economics literature on discrimination in order to determine: (1) how often papers seek to distinguish between taste-based and belief-based (statistical) sources of discrimination; (2) how often papers seek to distinguish between accurate and inaccurate beliefs for belief-based sources of discrimination. The methodology and inclusion criteria are outlined in the Supplemental Material. It is important to note that while this section covers the empirical literature, the potential for inaccurate beliefs has also been discussed in theoretical discrimination research (Arrow, 1973, 1998; Schwartzstein, 2014).[10]

Table 1 tabulates the 105 papers published in 10 top economics journals between 1990 and 2018 that test for evidence of discrimination. Most papers that met our inclusion criteria found evidence of discrimination: 102 out of 105 papers, or 97.1% documented evidence for discrimination against at least one group that was considered in the paper. The majority of papers (61.9%) discussed the source of discrimination as being driven by either preferences (taste-based) or beliefs (statistical), and nearly half of the papers (46.7%) attempted to distinguish between these two sources through a formal test. However, very few papers even discussed the possibility that beliefs may be inaccurate (10.5%), and fewer still examined whether beliefs were accurate or inaccurate (6.7%).[11] Despite the lack of discussion and explicit tests, we would argue that inaccurate statistical discrimination is a reasonable alternative interpretation in nearly all of these

[10] Arrow (1973) notes that employers may be more willing to accept subjective probabilities that accord with their actions. Similarly, Arrow (1998) writes "the discussion of statistical discrimination so far assumes that the employers or creditors use all the information available throughout the economy. In Bayesian terms, the posterior information is sufficiently rich to make the contribution of the prior minimal. But of course this is not so." Notably, in neither paper is the potential of inaccurate beliefs addressed formally.

[11] The papers that tested for inaccurate beliefs include Fershtman and Gneezy (2001); List (2004); Mobius and Rosenblat (2006); Beaman, Chattopadhyay, Duflo, Pande, and Topalova (2009); Agan and Starr (2017); Hedegaard and Tyran (2018); Arnold et al. (2018). We discuss the methods and findings of these papers further in the Supplemental Material.

Table 1. Summary of Literature Review on Discrimination

	All: 1990 - 2018		**Recent: 2014 - 2018**	
	# Papers	*% Total*	*# Papers*	*% Total*
Papers meeting inclusion criteria	105	100.0%	31	100.0%
Evidence of discrimination	102	97.1%	31	100.0%
Discuss taste-based versus statistical source	65	61.9%	23	74.2%
Test for taste-based versus statistical source	49	46.7%	16	51.6%
Discuss accurate versus inaccurate beliefs	11	10.5%	5	16.1%
Test for inaccurate beliefs	7	6.7%	3	9.7%
Measure beliefs	7	6.7%	3	9.7%

cases.

3 A Model of Discrimination with Inaccurate Beliefs

In this section, we model discrimination with inaccurate beliefs in the context of a simple hiring decision. An evaluator learns about a worker's productivity from a signal, then decides whether to hire the worker. Inaccurate beliefs refer to the case where the evaluator misperceives either how the distribution of productivity or the signal distribution varies by group identity. We use this model to explore how a researcher can identify the source of discrimination. We first show that many different preferences and beliefs generate an identical pattern of discrimination, creating an identification challenge. We then show that, when allowing for inaccurate beliefs, the commonly used outcomes-based method can reject the possibility of accurate statistical discrimination but it cannot separate whether discrimination stems from preferences or inaccurate beliefs. Further, erroneously assuming accurate beliefs and using the outcomes-based method leads to a misclassification of discrimination driven by inaccurate beliefs as arising from preferences. We conclude by outlining two alternative methods—eliciting beliefs and manipulating information—to identify whether discrimination stems from preferences or beliefs. A reader who prefers to skip the formal presentation of the theory can jump to the empirics in Section 4.

3.1 Model

Workers. Consider a worker who has observable group identity $g \in \{M, F\}$ and unobservable productivity a drawn from normal distribution $N(\mu_g, 1/\tau_g)$, with mean productivity $\mu_g \in \mathbb{R}$ and concentration of productivity $\tau_g > 0$. The worker completes a task,

such as an interview or test, that generates a signal of productivity $s = a + \epsilon$, where $\epsilon \sim N(0, 1/\eta_g)$ with signal precision $\eta_g > 0$.[12] Without loss of generality, we focus on discrimination against workers from group F.

Evaluators. An evaluator decides whether to hire the worker, $v \in \{0, 1\}$ where 1 corresponds to hire and 0 corresponds to do not hire. Before making this decision, the evaluator observes the worker's group identity g and realized signal s.

We model inaccurate beliefs as a misspecified model of the group-specific productivity and signal distributions. Namely, the evaluator holds subjective beliefs $\hat{\mu}_g \in \mathbb{R}$ and $\hat{\tau}_g > 0$ about the mean and concentration of productivity for group g, and subjective belief $\hat{\eta}_g > 0$ about the precision of the signal for group g. Inaccurate beliefs corresponds to the case in which these subjective distributions differ from the true distributions.[13] The evaluator uses Bayes rule with respect to these subjective distributions to form a posterior belief about the worker's productivity. From Bohren and Hauser (2021), this misspecified model framework can capture a variety of biases and heuristics in belief formation that have been documented in the literature, including non-Bayesian updating rules.

The evaluator hires the worker if her subjective posterior belief about expected productivity is above a group-specific hiring threshold $u_g \in \mathbb{R}$. This threshold is a reduced form representation of how the evaluator's preferences depend on productivity and group identity.[14] We refer to the evaluator's preferences and subjective beliefs as her type, denoted by $\theta \equiv (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g \in \{M,F\}}$. Let $v(s, g, \theta) \equiv \mathbb{1}\{\hat{E}_\theta[a|s, g] \geq u_g\}$ denote the optimal hiring decision by an evaluator of type θ who observes a worker from group g with realized signal s, where $\hat{E}_\theta$ denotes the expectation taken with respect to θ's subjective beliefs.

We next categorize different forms of preferences and beliefs. We use the term *par-*

[12] To distinguish the two variance parameters, we refer to τ_g as the concentration of productivity and to η_g as the signal precision.

[13] An additional form of inaccurate beliefs that we do not consider is the possibility that an evaluator believes that the mean of the signal differs by group identity. For example, all signals for group F are inflated by a constant $b > 0$ i.e. $s = a + b + \epsilon$, and therefore, the evaluator discounts a signal to $s - b$ for group F.

[14] The microfoundation for this reduced form is as follows. If the evaluator hires the worker, she earns a payoff that is linear in productivity and also depends on group identity, $m_g a + b_g$, where $m_g > 0$ is a group-specific marginal value of productivity and $b_g \in \mathbb{R}$ is a group-specific taste parameter. If she does not hire the worker, she earns outside option $\underline{u}$. The evaluator maximizes her expected payoff. She hires the worker if and only if $\hat{E}[m_g a + b_g|s, g] > \underline{u}$, or $\hat{E}[a|s, g] > (\underline{u} - b_g)/m_g \equiv u_g$, where $\hat{E}$ denotes the expectation with respect to the evaluator's subjective beliefs. Therefore, u_g is a reduced form representation of the evaluator's payoff.

tiality to refer to properties of these model primitives. An evaluator with *preference partiality* sets different expected productivity thresholds for hiring workers from groups F and M.

Definition 1 (Preference Partiality)**.** *An evaluator has preference partiality against group F if $u_F > u_M$, preference partiality against group M if $u_M > u_F$, and preference neutrality if $u_F = u_M$.*

Preference partiality leads the evaluator to make different hiring decisions when she has the same posterior belief about the expected productivity of a worker from each group. An evaluator with *belief partiality* has different subjective beliefs about the productivity and/or signal distributions for each group.

Definition 2 (Belief Partiality)**.** *An evaluator has* belief partiality *if $(\hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) \neq (\hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M)$ and* belief neutrality *if $(\hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) = (\hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M)$. This belief partiality stems from (i) lower expected productivity if $\hat{\mu}_F < \hat{\mu}_M$; (ii) lower (higher) concentration if $\hat{\tau}_F < \hat{\tau}_M$ ($\hat{\tau}_F > \hat{\tau}_M$); and (iii) lower (higher) signal precision if $\hat{\eta}_F < \hat{\eta}_M$ ($\hat{\eta}_F > \hat{\eta}_M$). Belief partiality is* accurate *if $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) = (\mu_g, \tau_g, \eta_g)$ for $g \in \{M, F\}$ and otherwise is* inaccurate*.*

Discrimination. Following the definition proposed in Bohren et al. (2022), we focus on *direct* discrimination, which is based on the difference in the hiring decision for a worker from group M versus F with the same realized signal. Let

$$D(s, \theta) \equiv v(s, M, \theta) - v(s, F, \theta) \tag{1}$$

denote this difference for an evaluator of type θ who observes realized signal s. *Direct discrimination* occurs at s when $D(s, \theta) \neq 0$; it occurs against group F if $D(s, \theta) > 0$ and against group M if $D(s, \theta) < 0$. There is *no discrimination* if $D(s, \theta) = 0$ for all $s \in \mathbb{R}$. When different sets of beliefs and preferences give rise to the same discriminatory behavior at all signals, we refer to this as *equivalent discrimination*.

Definition 3 (Equivalent Discrimination)**.** *Two evaluators of types θ and θ' exhibit* equivalent discrimination *if $D(s, \theta) = D(s, \theta')$ for all $s \in \mathbb{R}$.*

While partiality refers to evaluators' preferences and beliefs, *discrimination* is a property of behavior and a consequence of these primitives. Identifying the source of discrimination refers to determining which form(s) of partiality generate the observed

discrimination. Using this terminology, what the literature often refers to as taste-based discrimination corresponds to differential treatment stemming from preference partiality, while what is often referred to as statistical discrimination corresponds to differential treatment stemming from belief partiality. We define *inaccurate statistical discrimination* as differential treatment stemming from inaccurate belief partiality.

Discussion of Model. We focus on binary evaluations for a population of workers with normally distributed productivity and signals. This simple set-up allows us to illustrate how inaccurate beliefs impact discrimination in a tractable and succinct way. Our set-up easily extends to alternative forms of evaluations—e.g. selecting a wage offer as in the experimental setting considered in Section 4 or a rating from a non-binary discrete set—and to other productivity and signal distributions.

In terms of identifying the source of discrimination, we focus on sources of *direct* discrimination, where workers from different groups receive differential treatment conditional on generating the same information. A broader definition of discrimination, termed *total discrimination*, considers differential treatment conditional on underlying productivity a or some other qualification. This broader definition encompasses both direct and indirect, or *systemic*, discrimination (Bohren et al., 2022). While it is likely that inaccurate beliefs present an identification challenge for identifying the source of systemic discrimination as well, a formal analysis of this is beyond the scope of the current paper.

3.2 Optimal Hiring Rule and Equivalent Discrimination

We next derive how the optimal hiring rule depends on preferences and beliefs. Given signal s and group identity g, the evaluator's posterior belief about productivity is normally distributed with mean $\hat{\mu}_g(s,\theta) \equiv (\hat{\tau}_g\hat{\mu}_g + \hat{\eta}_g s)/(\hat{\tau}_g + \hat{\eta}_g)$ and variance $1/(\hat{\tau}_g + \hat{\eta}_g)$. Since the posterior mean is monotonic with respect to s, the optimal hiring rule can be represented as a cut-off with respect to the signal.

Lemma 1 (Optimal Hiring Rule)**.** *A type θ evaluator hires a worker from group g who generates signal s, i.e. $v(s,g,\theta) = 1$, if and only the signal is weakly greater than*

$$\overline{s}(\theta, g) \equiv \left(\frac{\hat{\tau}_g + \hat{\eta}_g}{\hat{\eta}_g}\right) u_g - \frac{\hat{\tau}_g}{\hat{\eta}_g}\hat{\mu}_g. \tag{2}$$

The signal required to hire a worker is increasing in the evaluator's preference u_g and decreasing in the prior belief about average productivity $\hat{\mu}_g$. When $\hat{\mu}_g < u_g$, it is increas-

ing in the concentration of productivity $\hat{\tau}_g$ and decreasing in the signal precision $\hat{\eta}_g$. In this case, the evaluator seeks workers perceived to be in the top tail of the productivity distribution. Therefore, a higher signal realization is required to offset a concentrated productivity distribution. In contrast, the evaluator is willing to hire at lower signal realizations when the signal is more precise. These comparative statics reverse when $\hat{\mu}_g > u_g$ and the evaluator seeks to avoid workers perceived to be in the bottom tail.

We use Lemma 1 to derive the sets of beliefs and preferences that give rise to equivalent discrimination. An evaluator of type θ discriminates against group F if she sets a higher hiring rule for group F, $\overline{s}(\theta, F) > \overline{s}(\theta, M)$. Types exhibit equivalent discrimination when they have preferences and beliefs that lead to the same pair of hiring rules.

Lemma 2 (Equivalent Discrimination)**.** *For any constants* $(s_M, s_F) \in \mathbb{R}^2$ *with* $s_F > s_M$*, the set of types*

$$\{\theta \mid \overline{s}(\theta, M) = s_M \text{ and } \overline{s}(\theta, F) = s_F\} \tag{3}$$

exhibit equivalent discrimination against group F*. For each* $(s_M, s_F) \in \mathbb{R}^2$ *such that* $s_M = s_F$*, the set of types that satisfy Eq. (3) exhibit no discrimination.*

We can represent the sets of types that exhibit equivalent discrimination as a pair of level sets parameterized by $(s_M, s_F) \in \mathbb{R}^2$, which we refer to as an *isodiscrimination curve*.

Fig. 1.a illustrates an isodiscrimination curve in two-dimensions. Fixing the other parameters, it plots the continuum of preference parameters and subjective average productivities for group F that lead to a given pair of hiring rules. For example, an evaluator with mild preference partiality and extreme belief partiality exhibits equivalent discrimination to an evaluator with more extreme preference partiality and mild belief partiality. The blue dotted line traces out preference neutrality (i.e. $u_F = u_M$) and the green dotted line traces out belief neutrality (i.e. $\hat{\mu}_F = \hat{\mu}_M$). As can be seen in the figure, a given pattern of discrimination can stem from both preference and belief partiality against group F, belief partiality that is somewhat offset by more favorable preferences, or vice versa.

3.3 Identifying the Source of Discrimination.

Researchers are often interested in identifying the *source* of discrimination i.e. the form of partiality that generates the observed discriminatory behavior.[15] Manski (2004) first

[15] A property is *identified* if it can be backed out from available data, or more formally, if there exists an injective relationship between the observed data and the property (Haavelmo, 1944).

observed that choice behavior could be consistent with multiple sets of preferences and subjective beliefs, and hence, lead to difficulties when using choice data to identify preferences and beliefs. We explore how the possibility of inaccurate beliefs impacts such identification in relation to discrimination.

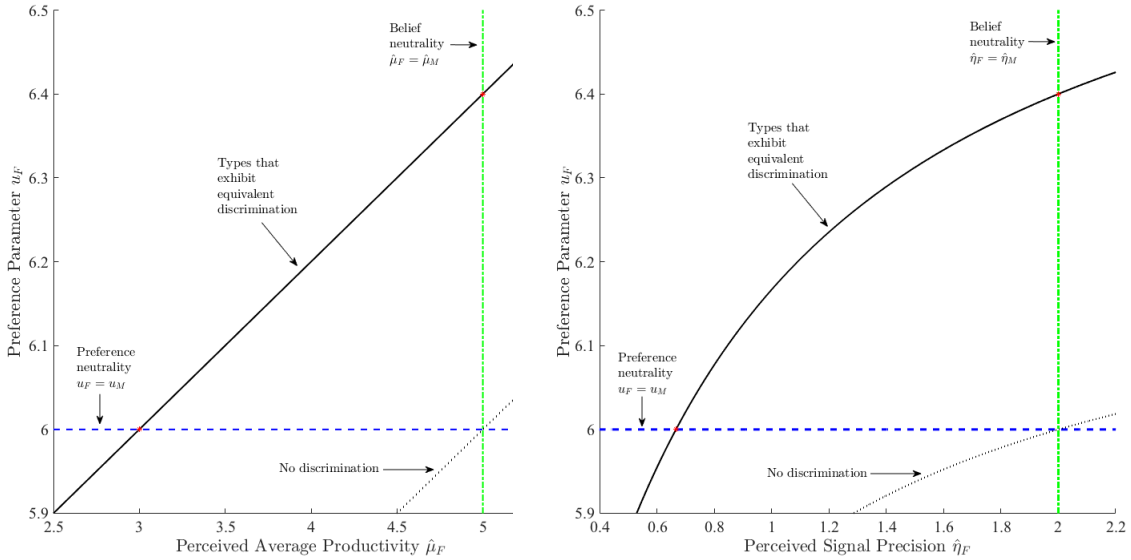
To proceed, we assume that the researcher observes the group identity g, realized signal s and hiring decision v for each worker, and that the dataset includes a sufficiently rich set of workers such that the hiring rule for each group can be identified from this data—that is, the pair of signal cut-offs, which we denote by $(s_M, s_F) \in \mathbb{R}^2$.[16]

An Identification Challenge. It is well known that measuring discrimination (e.g. the extent to which workers who generate similar signals receive different evaluations) cannot be used to distinguish between preference-based partiality and accurate belief-based partiality about the productivity distribution (see for example Bertrand and Mullainathan 2004). The same insight extends to inaccurate beliefs about the distribution of productivity and accurate or inaccurate beliefs about the signal distribution. To formalize this insight, we show that for any pair of hiring rules, each form of partiality in isolation can generate the given pattern of discrimination. Therefore, observing the hiring rule for each group does not rule out either preference-based partiality or any of the forms of belief-based partiality.

Proposition 1 (Equivalent Sources)**.** *For any pair of hiring rules* $(s_M, s_F) \in \mathbb{R}^2$ *with* $s_F > s_M$*, a continuum of types exhibit equivalent discrimination, including:*

1. *A type with* preference partiality *and belief neutrality,* $u_F > u_M$ *and* $(\hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) = (\hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M)$*;*
2. *A type with preference neutrality and* belief partiality due to lower expected productivity*,* $\hat{\mu}_F < \hat{\mu}_M$ *and* $(u_F, \hat{\tau}_F, \hat{\eta}_F) = (u_M, \hat{\tau}_M, \hat{\eta}_M)$*;*
3. *A type with preference neutrality and* belief partiality due to higher concentration of productivity*,* $\hat{\tau}_F > \hat{\tau}_M$ *and* $(u_F, \hat{\mu}_F, \hat{\eta}_F) = (u_M, \hat{\mu}_M, \hat{\eta}_M)$*, and also such a type with belief partiality due to lower concentration of productivity;*
4. *A type with preference neutrality and* belief partiality due to higher signal precision*,*

[16] In practice, observing signals directly may not be possible. An alternative method is a *correspondence study*, which randomly assigns group identity and signals to a set of fictitious workers, then elicits hiring decisions (for example, the classic resume study of Bertrand and Mullainathan (2004)). This ensures that workers from each group in the fictitious sample have the same distribution over signals, and therefore, any differences in hiring can be causally attributed to group identity. An audit study uses a similar randomized procedure to identify discrimination—experimental confederates with different group identities interact with evaluators while following the same script.



(a) Preference versus subjective average productivity; $(\hat{\tau}_F, \hat{\eta}_F) = (.5, 2)$

(b) Preference versus subjective signal precision; $(\hat{\mu}_F, \hat{\tau}_F) = (5, .5)$

Figure 1. Equivalent Sources

Hiring rules $(s_M, s_F) = (6.25, 6.75)$; $(u_M, \hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M) = (6, 5, .5, 2)$; red asterisks denote types with single form of partiality from Proposition 1.

$\hat{\eta}_F > \hat{\eta}_M$ and $(u_F, \hat{\mu}_F, \hat{\tau}_F) = (u_M, \hat{\mu}_M, \hat{\tau}_M)$, and also such a type with belief partiality due to lower signal precision.

From Proposition 1, when all other parameters are equal, a higher preference parameter or a lower subjective average productivity for group F relative to M generates discrimination against group F. For the other parameters, all else equal, a higher subjective concentration of productivity or a lower subjective signal precision for group F generate discrimination against F for types with $\hat{\mu}_g < u_g$, while the opposite holds for types with $\hat{\mu}_g > u_g$. This stems from the comparative static in Eq. (2) for the variance parameter of interest: as discussed following Lemma 1, how the parameter impacts the signal thresholds, and therefore, the pattern of discrimination, depends on $\hat{\mu}_g$ and u_g.

Fig. 1 illustrates the evaluator types constructed in Proposition 1. Fixing a pair of hiring rules and holding the concentration and precision parameters equal across groups, Panel (a) illustrates the preference parameters and subjective average productivities for group F that lead to equivalent discrimination. This includes the type described in part (i), denoted by the asterisk where the isodiscrimination curve intersects the line of

belief neutrality, and the type described in part (ii), denoted by the asterisk where the isodiscrimination curve intersects the line of preference neutrality. Panel (b) repeats this exercise for the preference parameters and subjective signal precisions, illustrating the types described in parts (i) and (iv).

We next discuss methods that seek to separate preference and belief-based sources.

Outcomes-Based Test. A common method used to identify the source of discrimination under the assumption of accurate beliefs is to compare evaluations to the outcome distribution for each group. In the current framework, the outcomes-based test corresponds to comparing hiring rules to the true productivity and signal distributions. Clearly this requires the researcher to identify the true productivity and signal distributions, in addition to the hiring rules:

Suppose the researcher can identify the hiring rules $(s_M, s_F) \in \mathbb{R}^2$ *and the true productivity and signal distributions* (μ_g, τ_g, η_g) *for each group* $g \in \{M, F\}$.

Under the assumption of accurate beliefs, the outcomes-based test identifies the evaluator's preference parameters (u_F, u_M), and therefore, the source(s) of discrimination.[17] This is illustrated in Fig. 2: the unique preference parameters that are consistent with the observed hiring rules and true distributions are $u_F = 6.3$ and $u_M = 6$. Since $\mu_F = 4.5$ and $\mu_M = 5$, this evaluator has both preference partiality and accurate belief partiality.

We next explore how erroneously assuming accurate beliefs impacts the conclusions that a researcher draws from an outcomes-based test. To do so, we first define how inaccurate beliefs impact discrimination. We say a type's inaccurate beliefs increase discrimination if the type sets a higher hiring rule for group F and a lower hiring rule for group M, relative to the type with accurate beliefs and the same preferences, and similarly for decreasing discrimination.

Definition 4 (Increasing and Decreasing Discrimination)**.** *Suppose type* θ^* *has accurate beliefs and type* θ *has the same preferences as* θ^* *but inaccurate beliefs. Then* θ*'s inaccurate beliefs* increase *discrimination against group* F *if, relative to* θ^*, $\overline{s}(\theta, F) \geq \overline{s}(\theta^*, F)$ *and* $\overline{s}(\theta, M) \leq \overline{s}(\theta^*, M)$*, with one inequality strict. The definition of decreasing discrimination is analogous with the inequalities reversed.*

The validity of the accurate beliefs assumption is crucial: we next show that when the researcher erroneously assumes accurate beliefs and uses an outcomes-based test, she

[17] See Lemma 3 in Appendix A for a formal statement of this insight.

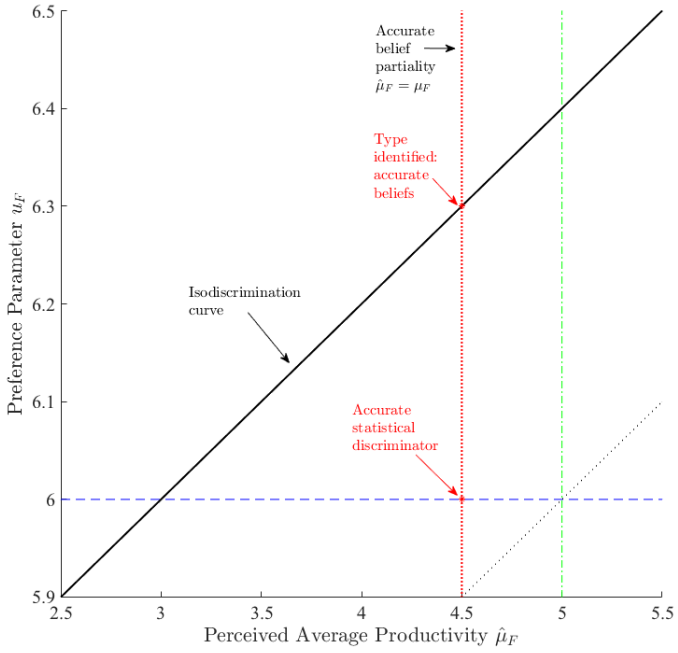


Figure 2. Outcomes-based Test
Isodiscrimination curve $(s_M, s_F) = (6.25, 6.75)$; true distributions $(\mu_M, \tau_M, \eta_M) = (5, .5, 2)$, $(\mu_F, \tau_F, \eta_F) = (4.5, .5, 2)$; $(u_M, \hat{\mu}_M, \hat{\tau}_M, \hat{\eta}_M) = (6, 5, .5, 2)$; blue dotted line: preference neutrality $(u_F = u_M)$; green dotted line: belief neutrality $(\hat{\mu}_F = \hat{\mu}_M)$.

mistakenly attributes discrimination stemming from inaccurate beliefs to a preference-based source. Depending on whether the inaccurate beliefs increase or decrease discrimination, the misidentified preference parameters will over- or underestimate the level of preference partiality.

Proposition 2 (Misclassification of Source)**.** *Suppose a researcher identifies the hiring rules* (s_M, s_F) *and true distributions* (μ_g, τ_g, η_g) *for* $g \in \{M, F\}$*. If a researcher incorrectly assumes an evaluator has accurate beliefs and uses an outcomes-based test to identify the source of discrimination, then for a generic set of types and true distributions, the researcher misidentifies the evaluator's type. If inaccurate beliefs* increase *discrimination against group* F*, then the researcher overestimates the evaluator's preference partiality against group* F*, while if inaccurate beliefs* decrease *discrimination, then the researcher underestimates preference partiality.*

Fig. 2 illustrates this result. If the evaluator believes that the average productivity for group F is $\hat{\mu}_F = 3$ when in fact it is $\mu_F = 4.5$, then incorrectly assuming accurate

beliefs will lead a researcher to conclude that $u_F = 6.3$ and $u_M = 6$, when in actuality, the evaluator has preference neutrality, $u_F = u_M = 6$. Therefore, the researcher attributes discrimination stemming from this inaccurate belief about average productivity to preference partiality.

While erroneously assuming accurate beliefs leads to a misclassification of source, the outcomes-based test no longer identifies the source when one relaxes this assumption. From Eq. (2), it is clear that when beliefs may be inaccurate, identifying the true belief parameters does not identify the evaluator's preferences. It can only be used to potentially rule out *accurate statistical discrimination*—that is, discrimination stemming from accurate beliefs and preference neutrality. The following result establishes when the observed pattern of discrimination is inconsistent with accurate statistical discrimination.

Proposition 3 (Rejecting Accurate Statistical Discrimination)**.** *Suppose a researcher identifies the hiring rules* (s_M, s_F) *and true distributions* (μ_g, τ_g, η_g) *for* $g \in \{M, F\}$*. If*

$$\frac{\tau_M \mu_M + \eta_M s_M}{\tau_M + \eta_M} \neq \frac{\tau_F \mu_F + \eta_F s_F}{\tau_F + \eta_F}, \tag{4}$$

then the evaluator is not an accurate statistical discriminator.

Accurate statistical discrimination is of particular interest because it is often viewed as efficient from an informational perspective and has been used to justify social stereotyping and rationalize discriminatory behavior (Tilcsik, 2021).[18] When Eq. (4) holds, the researcher can reject this explanation and conclude that the observed discrimination either stems from animus towards a group or inaccurate beliefs about them. In Fig. 2, the accurate statistical discriminator type does not lie on the isodiscrimination curve, and therefore, is not consistent with the observed hiring rules.[19]

[18]The main argument is that, if an evaluator is applying differential treatment to groups when underlying differences do exist, then this evaluator is simply using information in an optimal way and engaging in profit-maximizing behavior.

[19]Prior work has highlighted additional identification challenges for outcomes-based tests, including the problems of infra-marginality (Ayres, 2002; Simoiu, Corbett-Davies, and Goel, 2017) and relying on administrative data that may condition on a post-treatment outcome (Knox, Lowe, and Mummolo, 2020). In contemporaneous theoretical work, Hull (2021) shows that outcome-based tests that use IV and MTE methods (e.g. Arnold et al. (2018); Grau and Vergara (2021)) can distinguish between accurate statistical discrimination and other sources. These marginal outcome tests do not require the researcher to observe the decision-maker's signal and do not suffer from inframarginality or selection problems by definition. However, they still require additional assumptions in order to separate taste-based discrimination from inaccurate beliefs (see, for example, the structural model developed in Arnold et al. (2018)).

Of course, when the observed pattern of discrimination is consistent with accurate statistical discrimination, this does not identify the evaluator as an accurate statistical discriminator: other preferences and beliefs can also generate the observed behavior. Even in this case, it is still important to identify the source of discrimination: while a type with inaccurate beliefs may exhibit equivalent behavior to an accurate statistical discriminator for the current hiring decision, these inaccurate beliefs may affect the worker in future performance evaluations and promotions in ways that differ from accurate beliefs. For example, consider an evaluator who overestimates the difference in average productivity between groups and has preferences that somewhat favor the disadvantaged group for entry-level positions. Suppose this yields equivalent discrimination to an accurate statistical discriminator. Then if the evaluator only feels compelled to favor the disadvantaged group for entry-level hiring, these inaccurate beliefs will lead to persistently lower rates of promotion and advancement for the disadvantaged group.

Given the difficulty of using an outcomes-based test to identify the source of discrimination, we next explore two alternative methods.

Eliciting Beliefs. If it is possible to collect data on the evaluator's subjective beliefs, then comparing hiring decisions to these beliefs can identify the source of discrimination.[20]

Suppose the researcher can identify the hiring rules $(s_M, s_F) \in \mathbb{R}^2$ *and the subjective productivity and signal distributions* $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$ *for each group* $g \in \{M, F\}$.

One way to identify subjective beliefs would be to directly elicit them from evaluators.

Similar to the outcomes-based test under the assumption of accurate beliefs, this method identifies the evaluator's preferences, and therefore, the source of discrimination.

Proposition 4 (Identifying Preferences from Subjective Beliefs)**.** *Suppose a researcher identifies the hiring rules* (s_M, s_F) *and subjective beliefs* $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$ *for* $g \in \{M, F\}$*. This identifies the preference parameters* (u_M, u_F)*, and therefore, the evaluator's type.*

Importantly, observing subjective beliefs does not identify whether they are accurate—additional data, such as outcomes, is necessary to determine this.[21]

[20] Manski (2004) first proposed combining data on expectations with choice data to identify preferences without assuming rational expectations.

[21] An alternative methodology involves eliciting beliefs about group performance and comparing evaluations when the same groups are identified either using labels subject to stereotypes (e.g. gender) or not (e.g. birth month) (Coffman, Exley, and Niederle, 2021). Since the performance distributions

In practice, this method will be difficult in many settings—both due to the complexity and reliability of methods for eliciting beliefs about higher moments and due to the feasibility of collecting such information (for example, it may not be possible to collect beliefs in certain settings such as on an online platform). The next method provides an alternative, simpler way to partially identify the source of discrimination.

Manipulating Information. Suppose it is possible to manipulate the amount of information presented to evaluators. For example, one could compare discrimination in a treatment in which only one customer review is revealed to a treatment in which five customer reviews are revealed. In the current framework, we model this as varying the number of signal draws x that the evaluator observes for a worker.

Suppose the researcher can identify the hiring rules $(s_M^i, s_F^i) \in \mathbb{R}^2$ *for multiple informational treatments* i *with* x_i *signal draws.*

If an evaluator believes that a single draw of the signal has precision $\hat{\eta}_g$, then she believes that observing $x \geq 1$ conditionally independent draws of this signal has precision $x\hat{\eta}_g$. The characterization of the optimal hiring rules and set of types that generate equivalent discrimination following x draws is identical to the case of a single draw, substituting $x\hat{\eta}_g$ for $\hat{\eta}_g$.

We next establish that manipulating the number of signal draws can separate preference partiality from belief partiality, but it cannot separate the different forms of belief partiality. Proposition 5 establishes that for any two informational treatments, there is a unique pair of preference parameters that yield equivalent discrimination. However, there are a continuum of types with the same pair of preference parameters and distinct belief parameters that exhibit equivalent discrimination across both informational treatments. Moreover, this set of types also exhibit equivalent discrimination across all informational treatments. Therefore, identifying the hiring rules for at least two informational treatments identifies the evaluator's preferences u_g, but does not identify beliefs $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$.

Proposition 5 (Identifying Preferences from Manipulating Information)**.** *Suppose a researcher identifies the hiring rules* (s_M, s_F) *and* (s'_M, s'_F) *for two informational treatments corresponding to an evaluator observing either* $x \geq 1$ *or* $x' \neq x$ *signal draws for*

are the same regardless of the label, any differences in evaluations between the two treatments can be assigned to tastes rather than beliefs. As the authors note, creating equivalent evaluation settings for both types of labels requires that the methodology be implemented in a controlled laboratory environment.

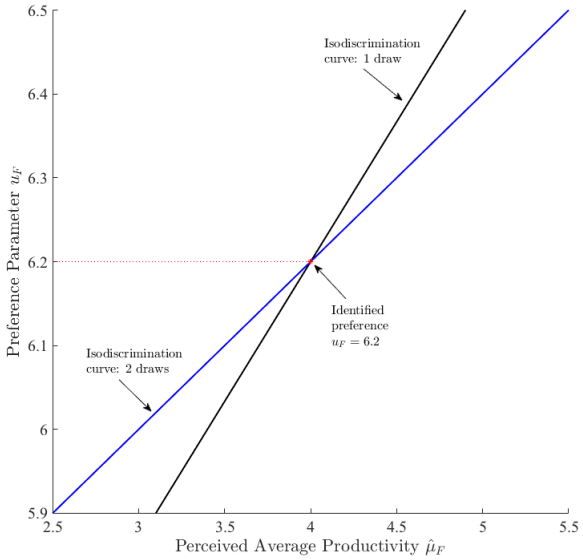


Figure 3. Information Manipulation
Isodiscrimination curves for $(u_F, \hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) = (6.2, 4, .5, 1)$

each worker. This identifies the preference parameters,

$$u_g = \frac{xs_g - x's'_g}{x - x'} \tag{5}$$

for $g \in \{M, F\}$, but does not identify beliefs $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$. Additional informational treatments provide no further identification of beliefs.

Fig. 3 illustrates Proposition 5. Only types with preference parameter $u_F = 6.2$ exhibit the observed discrimination for both informational treatments. The types with other preference parameters on the isodiscrimination curve for one signal draw do not exhibit equivalent discrimination when there are two signal draws—as can be seen in the figure, they do not also lie on the isodiscrimination curve for two draws.[22]

A crucial requirement for this result is that multiple signals are drawn from the same distribution. This ensures that the evaluator has the same belief about the signal distribution for each draw. One common real world setting that satisfies this requirement is

[22]While it may look like $\hat{\mu}_F$ is also identified as 4, this is just the belief for a type with $\hat{\tau}_F = .5$ and $\hat{\eta}_F = 1$. There are other types with different values of $\hat{\tau}_F$ and $\hat{\eta}_F$ that exhibit equivalent discrimination across one and two draws, and these types will have different values of $\hat{\mu}_F$ (but the same preference parameter $u_F = 6.2$). For example, from Eq. (16) in Appendix A, a type with $(u_F, \hat{\mu}'_F, \hat{\tau}'_F, \hat{\eta}'_F) = (6.2, 3.45, .4, 1)$ will exhibit equivalent discrimination to $(u_F, \hat{\mu}_F, \hat{\tau}_F, \hat{\eta}_F) = (6.2, 4, .5, 1)$ across one and two draws.

when each signal draw is a past review or rating, such as on Airbnb, and the worker is rated multiple times by evaluators from the same population. A researcher could manipulate information by varying the number of ratings that are visible to evaluators. Another example is settings where a worker receives multiple simultaneous recommendations that serve as signals, such as reviews for a grant proposal or recommendations for employment from colleagues with similar qualifications. In this case, a researcher could manipulate information by varying the number of recommendations that are shown to the evaluating committee. In contrast, varying observation of signals from different domains—for example, comparing discrimination when education is observed to discrimination when education and SAT score are observed—does not identify the evaluator's preference partiality. In this case, the evaluator may have a different subjective signal distribution for the signal from each domain, and therefore, the result no longer holds.

Taken together, the proposed belief elicitation and information manipulation methods can be used to separate preference and belief partiality. If it is possible to elicit an evaluator's beliefs for all parameters of the relevant distributions, then it is possible to fully identify the evaluator's type. If not, then the information manipulation method provides an alternative, simpler way to identify preferences and "aggregate" belief partiality—although it comes at the cost of not being able to separate the different ways that beliefs may be inaccurate.

4 Identifying the Source of Discrimination in a Hiring Experiment

In this section, we employ a stylized experimental setting to demonstrate the pitfalls of the identification problem outlined in the previous section—in particular, how assuming accurate beliefs can lead to erroneous conclusions about the source of discrimination—and to illustrate how the belief elicitation method outlined above can solve it. The experiment allows us to observe the actual distribution of productivity by group, and therefore, to perform the accounting exercise employed in outcomes-based tests and to also elicit beliefs about relevant characteristics. We show that average beliefs about productivity are incorrect, thereby violating the accurate beliefs assumption, and that ignoring these inaccurate beliefs leads to a false identification of the source of discrimination. We also demonstrate how providing information about the true group-specific average productivities can be used both to separate inaccurate beliefs from underlying animus and to correct these inaccurate beliefs. Participants adjust their behavior significantly in the direction of the information, suggesting that at least some of the observed

discrimination is driven by inaccurate beliefs rather than animus.

4.1 Experimental Design

In this section, we provide a summary of the pre-registered experimental design.[23] We recruited two samples of subjects on Amazon Mechanical Turk (participants) to complete either a work task (Survey 1) or a hiring task (Survey 2), which we next describe in detail.

Survey 1 (Work Task). We recruited 589 subjects from MTurk on February 23, 2018 for the first survey.[24] The survey was posted with the title "Math Questions and Demographics" and the description "A 20-minute task of answering math questions." We paid \$2 (i.e. a projected \$6/hour wage) and recruited a subject pool of 392 from the United States and 197 from India, all of whom had completed at least 500 prior tasks and had an 80% or higher approval rate for these tasks.[25] After starting the survey, subjects were informed that they would first answer demographic questions and then answer 50 multiple choice math questions. They were told that their performance would not affect their payment, and were asked not to use a calculator or any outside help, but just to do their best. This was followed by seven questions that provided the information used for their profiles in the second survey: favorite color, favorite movie, coffee vs. tea preference, age, gender, favorite subject in high school, and favorite sport. The math test included a mix of arithmetic (e.g. "$5 * 6 * 7 =$?"), algebra (e.g. "If $(y + 9) * (y^2 - 121) = 0$, then which of the following cannot be y?"), and more conceptual questions (e.g. "Which of the following is not a prime number?"). Finally, subjects were thanked for their participation and informed that they may receive a small bonus based on a different experiment, for reasons unrelated to their performance on the task. We describe the basis for such bonuses in the description of Survey 2.

The purpose of the first survey was to create a bank of "workers" who could be

[23]The experiment was pre-registered on AsPredicted (#8678). There are two minor differences between the pre-registration plan and the actual study. First, we pre-registered that we would recruit 400 US employers in the hiring task survey, but then decided to target an additional 200 Indian employers so that we could examine in-group/out-group evaluations. Second, we did not pre-register sample restrictions due to completing the task too quickly or slowly. We dropped 12 subjects in the work task survey and 5 in the hiring task survey due to these restrictions. The full surveys are in the Supplemental Material

[24]We received 604 responses in total, but dropped 12 responses that corresponded to the top 1% ($<$ 227 seconds) and bottom 1% ($>$ 3274 seconds) in terms of survey duration. Of the remaining 592 responses, we dropped 3 whose Qualtrics survey responses could not be matched to their MTurk records, leaving 589 final respondents.

[25]This geographic restriction is based on the addresses MTurkers used to register on Amazon. The survey was posted as two tasks on MTurk, with one only eligible for Indian workers and one only eligible for U.S. workers.

hired by the "employers" in the second survey. This novel design has several advantages over the existing paradigms for studying discrimination in the field. First, in contrast to correspondence studies, we did not employ deception at any point—all profiles shown to employers corresponded to actual workers who would in fact be paid as described in the following paragraph. However, similar to a correspondence study, we were able to control the information seen by an employer about a prospective worker by constructing worker profiles that included information that is ostensibly relevant for animus and/or beliefs about productivity (e.g. age, gender, and nationality), as well as other non-target information (e.g. tea preference). The non-target information ensures that the relevant demographics are not the only salient information provided to the employer (this mimics the additional – ostensibly less decision-relevant – information contained on a CV). Finally, instead of the coarse measures of discrimination used in many other studies (e.g. callback or stop rates), we elicit relatively continuous and precise measures of productivity and discrimination that are tightly linked. The downside to such a design is that the target characteristics and productivity may be correlated with the non-target information and thus may inform their decisions.[26]

Survey 2: We recruited 577 different MTurk subjects on February 26, 2018. We used the same hiring criteria as the first survey (392 from U.S., 185 from India, >= 80% approval rate).[27] The survey was posted with the title "20-Minute Survey about Decision-Making" and the description "20-Minute Survey about Decision-Making." We paid $2 (i.e. a projected $6/hour wage). Subjects were first asked to report their gender, age, and education level. Subjects were then presented with the first hiring task portion of the survey.

First Hiring Task. We informed subjects that we had previously paid other sub-

[26]While we chose items that we intended to be less informative for the task at hand, but not entirely irrelevant, favorite high school subject was in fact both relevant for performance (those who mentioned math performed roughly 3.2 points higher on the math test) and anticipated in wage offers (they were offered wages 5 cents higher on average). The other items were much less relevant for both scores and wage offers. Since these were open responses, we can't flexibly control for every aspect of the profile that could influence employer decisions, but in our wage regressions in the Appendix, we show that nationality and gender remain significant even after controlling for binary versions of the other profile attributes. Ultimately, while the design used in this stylized experiment has some advantages (e.g. no deception), by not randomizing demographic characteristics, we are capturing a broader "bundle of sticks" (Sen and Wasow, 2016) of these identity attributes than would be the case in a standard correspondence or audit study. Incentivized resume rating (Kessler, Low, and Sullivan, 2019) may be an appealing alternative design that has no deception but maintains the randomized demographics.

[27]We recruited 587 subjects in total, but dropped 7 whose surveys were completed in under 300 seconds and 3 whose stimuli (the profiles they evaluated) could not be matched to the first survey.

jects ("workers") to answer 50 math questions, showed them five examples of the math questions, and told them that, on average, participants answered 36.95 out of 50 questions correctly. They were then told that they would act as an employer and hire one of these workers by stating a wage (paid as a bonus to the worker). In return, they would receive a payment based on how many questions their hired worker answered correctly. This was followed by a more detailed description of the assignment. Each "employer" would view 20 profiles of potential workers and state the highest wage (between 0 to 50 cents) they were willing to pay to each worker. The employer would be paid 1 cent for each question answered correctly by the hired worker. We next described the mechanism (Becker-Degroot-Marschak) used to assign payment. We would randomly select a profile from the 20 potential workers. We would then draw a random number from 0 to 50. If the wage the employer stated for the worker was equal or greater than that number, then the worker would receive the random number as a bonus and the employer would receive a "profit" equal to the worker's performance minus the random number. If instead the employer stated a wage for the worker that was lower than the random number, then neither the worker nor the employer would receive a payment.

To ensure comprehension, we showed subjects an example profile (see Fig. 4) and stated wage. We gave examples of actual performance and randomly generated numbers that would produce positive profit, negative profit, and no hiring. Having highlighted the possibility of negative profit, we then noted that all employers would automatically be paid a \$0.50 bonus in addition to any money made through the hiring task, so that no employers would owe money. Finally, we ran a comprehension check with the same example profile, a specific wage (43), a random number (18), and an actual performance (10). We required the employer to correctly state how many cents they would have to pay the worker (18) and how many cents the employer would be paid before subtracting off the amount they would pay the worker (10).[28] Finally, employers were presented with a second wage (15), and answered the same questions. They were then presented with 20 profiles, each randomly selected with replacement from the bank of 589 profiles produced by the first survey.

[28]Entering an incorrect an answer would generate a pop-up with "Wrong Answer" and restrict the individual from moving to the next page.

Figure 4. Example Profile Used in First Hiring Task Description

Country:	United States
Gender:	Female
Age:	63
Favorite High School Subject:	English
Favorite Sport:	Gymnastics
Favorite Color:	Sea Green
Favorite Movie:	Overboard
Prefers Coffee/Tea:	Tea

Belief Elicitation Task. Next, subjects were randomly assigned to one of two different conditions: an incentivized or un-incentivized belief elicitation. Across both conditions, subjects were reminded that the full sample answered 36.95 out of 50 questions correctly. They were then asked to answer six questions of the form, "On average, how many math questions out of 50 do you think X answered correctly?" where X corresponded to the groups "women", "men", "people from the United States", "people from India", "people below or at the age of 33," and "people above the age of 33."[29] In the incentivized condition, prior to the six questions, subjects were told that they could earn a significant bonus for an accurate prediction. One of the six questions would be randomly selected and they would be paid \$5 minus their deviation from the question (bounded below by \$0). For example, if they answered 40 and the true average was 37, they would receive a \$2 bonus. Finally, they were asked to "please answer the questions as carefully as possible so that you can potentially win a large bonus."

Information Intervention & Second Hiring Task: After completing the belief elicitation, subjects were shown the correct answer for all six groups: women (35.28), men (38.32), people from the U.S. (37.14), people from India (36.58), people below or at the age of 33 (37.10), and people above the age of 33 (36.79). As discussed in Section 1, providing accurate information about group-level statistics is one potential method for

[29]We only elicited beliefs about the first moment of the performance distribution. While participants may also have inaccurate beliefs about other statistics, demonstrating a difference in subjective versus actual means is sufficient to falsify the assumption that beliefs are correct, which was the primary goal of the illustrative experiment. Eliciting other moments of the distribution, e.g. variance, is more complex for participants relative to eliciting the mean. Given the multiple stages in the study, we sought to keep the belief elicitation task as simple as possible in order to curtail potential confusion and minimize noise.

differentiating between inaccurate beliefs and "animus-driven" beliefs. While the former should shift in the direction of the information, the latter are unlikely to be moved because the errors are due to non-informational factors. Following this information, we stated, "Now that you have learned those facts, we would like you to work on 10 more profiles." We noted that, as in the first hiring task, we would randomly select one profile and a number, and pay bonus and wages accordingly (with an additional $0.50 automatic bonus to ensure no negative payments). After employers reviewed the 10 additional worker profiles, we thanked them for their participation, noted that we would calculate bonuses and pay them within a week, and allowed subjects the option to leave comments.

Summary Statistics. Appendix Table B1 provides summary statistics for the full sample of subjects that completed surveys 1 and 2 (Column (1)), as well as these statistics for each of the 6 demographic groups used in the second survey. On average, the work task (survey 1) took subjects 19 minutes to complete, while the hiring task took 23 minutes. There is variation in this timing across groups. Subjects from the U.S. took an average of 19 minutes to complete the hiring task, while subjects from India took 31.60 minutes; a difference also reflected in their median times (15.8 vs. 25.6). Another large difference between the U.S. and India samples is the average age of participants; the average Indian subject in the work task is approximately 8 years younger than the average American subject. This gap shrinks to 4 years for the hiring task. The Indian sample also skews more male than the U.S. sample (68.5% vs. 48.2% and 76.8% vs. 51.4% for survey 1 and 2, respectively) and is more likely to have a college education or above (90.3% vs. 56% in survey 2; the question was not asked in survey 1). While we primarily focus on simple comparisons between each demographic group, these observed differences motivate our use of multivariate regressions in robustness tests contained in the Appendix.

Connection to Theoretical Framework from Section 3. In the experiment, productivity a corresponds to the worker's performance on the math test. The experimental design simplifies the theoretical framework by eliminating the signal of productivity—evaluators observe group identity but no performance signal. It also has a richer action space: subjects choose a wage between 0 and 50 cents, whereas the theoretical framework is based on a binary hiring decision. Given the induced payoffs in the experiment, the optimal action depends on the subjective average productivity but not the subjective variance of productivity. The analysis from Section 3 easily extends to this alternative

action space and decision rule. Given that there is only scope for belief-based partiality due to differences in subjective average productivity, in analyzing the experimental data, we focus on comparing average wages to measure discrimination and we elicit beliefs about average productivity to determine whether beliefs are inaccurate.

4.2 Experimental Results

A necessary prerequisite to study the source of discrimination is to find a context and a population in which discrimination occurs. Ex-ante, it was not obvious that our stylized hiring experiment would satisfy this requirement. The employers knew that they were being observed as part of a research study and the relevant group information was represented abstractly (e.g. written text) rather than viscerally (e.g. a picture). All of these factors may attenuate the influence of animus.[30]

Despite these attenuating factors, we did find evidence of discrimination with respect to two out of three group identities: gender and nationality. Panel A of Table 2 presents the differences in average wages paid by employers to worker profiles from each group. With respect to gender, male profiles were paid on average 31.90 cents, while female profiles were paid 30.85 cents, a significant 3.4% difference ($p < 0.01$). With respect to nationality, profiles from India were favored, earning an average of 32.85 cents, while profiles from the U.S. earned 30.71 cents, a significant 7.0% difference ($p < 0.01$). Finally, there was no statistically significant evidence of age discrimination: subjects at or below age 33 were paid an average of 31.67 cents and those above age 33 were paid 31.14 cents, a 1.7% difference ($p = 0.17$). Table B2 demonstrates that these results are relatively similar in a multiple regression framework with employer fixed effects, though adding additional some profile characteristics does attenuate differences (notably, "favorite high school subject", which is both predictive of productivity and wages and correlated with gender and nationality).

To examine the possibility of in-group bias, we run similar regressions controlling for the employer belonging to the group of interest (e.g. female) and the interaction of the two indicators to measure in-group bias (see Table B3). We find that the interaction is insignificant for gender and marginally significant for nationality, although in the direction of favoring the out-group. For age, we find a significant interaction effect. This suggests that the null effect in Table B2 masks in-group bias by both older and younger employers. Antonovics and Knight (2009) use a similar set of regressions to

[30]For example, Bar and Zussman (2019) argue that a lack of interaction may attenuate the extent of taste-based discrimination in driving test examinations.

Table 2. Wages and "Productivities", by Employee Characteristics (Hiring Task 1)

	(1) Group 1	(2) Group 2	(3) Diff.	(4) p-val	(5) #Obs. G1	(6) #Obs. G2
Panel A: Employers' Wage WTP, by Employee Characteristics						
Gender (1 = Male , 2 = Female)	31.90	30.85	1.05	0.01	6,306	5,234
	(12.07)	(12.23)				
Country (1 = US , 2 = India)	30.71	32.85	-2.14	0.00	7,700	3,840
	(12.20)	(11.95)				
Age (1 = Under 33 , 2 = Over 33)	31.67	31.14	0.54	0.17	6,139	5,401
	(12.00)	(12.33)				
Panel B: Employee Productivity, by Employee Characteristics						
Gender (1 = Male , 2 = Female)	38.30	34.98	3.32	0.00	6,306	5,234
	(8.55)	(8.73)				
Country (1 = US , 2 = India)	37.01	36.36	0.65	0.41	7,700	3,840
	(8.93)	(8.49)				
Age (1 = Under 33 , 2 = Over 33)	36.96	36.60	0.37	0.63	6,139	5,401
	(8.62)	(8.98)				

Notes: Standard deviations in parentheses. One observation per worker-employer combination. Column (4) shows the p-value from a regression of the outcome on a dummy variable for group membership, with standard errors two-way clustered by employer and worker.

test for taste-based discrimination. This specification is motivated by the assumption that animus varies between groups (i.e. there is less animus toward one's in-group than out-group), but that beliefs are similar across groups (since they are taking a "standard model of statistical discrimination" as the benchmark and note that "these beliefs must be correct in equilibrium"). In Table B4, we test this assumption in our experimental environment. We find that beliefs about the gender performance gap are identical among both female and male employers. However, for nationality, we find significant differences. Americans hold beliefs that favor the out-group and Indians hold beliefs favoring the in-group—while both groups believe Indians will outperform Americans, the latter group predicts a larger gap

Having demonstrated moderate levels of discrimination in hiring, we now examine the "ground truth" in actual productivity differences between groups. The typical outcomes-based test of statistical discrimination requires mapping disparities between groups in the evaluators' relevant decision (e.g. the wages offered to employees) to disparities in an outcome in the evaluators' objective function (e.g. the employees' productivity).[31] In our

[31]Translating the two measures may require strong modeling assumptions (e.g. whether there is heterogeneity in the search costs faced by evaluators). For discussions of these assumptions in the context of the hit-rate tests, see Antonovics and Knight (2009); Dharmapala and Ross (2004); Anwar and Fang (2006).

context, this requires mapping disparities in the employers' stated wages to disparities in group-specific productivity differences, i.e. the number of questions answered correctly. The commonly used outcome method compares disparities in wages to disparities in performance to measure the relative role of (accurate) statistical versus taste-based discrimination (in the context of our framework, accurate belief-based versus preference-based partiality). For simplicity, we will refer to both disparities as measured in "points."

Panel B of Table 2 shows the average number of correct answers by each sub-group (see Fig. B1 for probability density functions). As shown in Panel A of Table 2, the gap in average wages for men and women was lower than the gap in average performance (1.05 points versus 3.32 points).[32] Therefore, if we used the standard outcome method to separate statistical and taste-based discrimination, we would conclude that the entire 1.05 point disparity in wages is due to (accurate) statistical discrimination—the remaining 2.27 point difference in performance would be attributed to taste-based discrimination against men. Turning to nationality-based discrimination, there was a wage gap of -2.14 points in favor of Indians, compared to a performance gap of 0.65 points in favor of Americans. Under the standard approach, we would conclude that the -2.14 point disparity in wages, when compared to the +0.65 point difference in performance, suggests taste-based discrimination against Americans.[33]

We now proceed to examine whether inaccurate beliefs can explain the disparities in compensation. As an initial check to see whether employers' decisions were guided by the elicited beliefs, we correlate wages with their beliefs about group-specific productivities. We find positive correlations for all six groups of workers (Female: 0.12, Male: 0.12, India: 0.15, U.S.: 0.12, Over 33: 0.12, Under 33: 0.10). Given that we elicited beliefs after the hiring task, it is possible that part of these correlations are due to rationalization (e.g. an individual first discriminates against women when setting wages, then chooses

[32]We calculate productivity differences using the full sample of profiles observed in hiring task 1. This is a weighted sample of the original population of 577 workers (since each of the 589 employers saw independent random samples of 20 of the 577 workers). Due to the random variation in the profiles observed, the group-level averages slightly differ from those found in Table B1. For example, the male-female performance gap is 3.04 points in Table B1 and 3.32 points in this weighted sample. Note that the averages in Table B1 are the basis for the informational intervention.

[33]While we document significant discrimination by gender (i.e. men are paid more than women), the outcome method reveals that the performance gap exceeds the pay gap. This leads to the conclusion that there is taste-based discrimination *against* men. While the literature often equates taste-based discrimination with animus or prejudice, this link may be inappropriate when discrimination manifests as an equalizing action. For example, people may be equalizing wages between two groups despite differences in productivity due to fairness concerns. We discuss the implications of this distinction further in the conclusion.

Table 3. Beliefs about Productivity by Employee Characteristics

	Group 1 (1)	**Group 2** (2)	**Diff.** (3)	**p-val** (4)
Gender (1=Male, 2=Female)	34.04	32.14	1.89	0.00
	(8.26)	(8.41)		
Country (1=US, 2=India)	32.08	34.80	-2.72	0.00
	(8.56)	(9.44)		
Age (1=Under 33, 2=Over 33)	33.41	31.57	1.84	0.00
	(8.97)	(9.00)		

Notes: Standard deviations in parentheses. One observation per employer combination. Column (4) shows the p-value from one-sample t-tests for the equality of columns (1) and (2). # Observations = 577.

beliefs to justify this decision), or audience effects (e.g. an individual falsely reports beliefs that justify the discriminatory decision to the experimenter). To test for this, we provided half of the employees with large incentives for belief accuracy. In Table B5, we show that beliefs are nearly identical across both incentive conditions, with none of the six comparisons being significantly different from one another. Together these findings suggest that the employers' group-specific performance predictions provide meaningful information about their true beliefs.

In Table 3, we present employer beliefs about the group-specific average performance, which can be compared directly to the actual group-specific performance reported in Table 2, Panel B. Predictions about performance are lower than actual performance for all six groups. This overall underestimation is consistent with risk aversion (recall that employers face the potential of a negative payment, taken from their $0.50 bonus, if they overestimate performance). Consistent with this, gaps in beliefs about performance are larger than gaps in wage payments. Using employers' actual beliefs to identify the source of discrimination leads to substantially different conclusions than the outcomes-based method outlined above. Looking at nationality, the wage gap is -2.14 points and the performance gap is +0.65 points; the gap in beliefs is -2.72 points. Thus, the *entire* wage gap can be explained by inaccurate beliefs. In contrast to the outcome method which infers taste-based discrimination in favor of Indian workers, the remaining 0.58 point difference between the belief and wage gaps suggests prejudice *against* them. Looking at gender, the wage gap is 1.05 points, the performance gap is 3.32 points, and the belief gap is 1.89 points. The majority of the wage gap can be explained by inaccurate beliefs: the residual attributed to preference-based sources shrinks from 2.27 to 0.84 points. Finally,

Table 4. Effect of Information: Difference-in-Differences by Hiring Task

	(1)	(2)	(3)	(4)	(5)
Post-Info	1.53***	1.60***	1.06***	1.97***	2.33***
	(0.31)	(0.27)	(0.31)	(0.39)	(0.34)
Female	-1.05***			-0.66*	-0.80**
	(0.38)			(0.37)	(0.33)
Female X Post-Info	-0.64*			-0.89**	-1.01***
	(0.38)			(0.38)	(0.29)
Indian		2.14***		2.01***	2.02***
		(0.41)		(0.43)	(0.38)
Indian X Post-Info		-1.07**		-1.20***	-1.65***
		(0.43)		(0.44)	(0.33)
Over 33			-0.54	0.06	0.29
			(0.39)	(0.39)	(0.35)
Over 33 X Post-Info			0.41	0.12	-0.21
			(0.42)	(0.42)	(0.31)
N	17,310	17,310	17,310	17,310	17,310
R^2	0.01	0.01	0.00	0.01	0.48
DepVarMean	31.90	30.71	31.67	30.71	30.71
Employer FE?	No	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses, two-way clustered by employer and worker. "DepVarMean" is the mean of the dependent variable (wage WTP) in the omitted group (e.g. Male Workers in Hiring Task 1 for column (1)). "Post-Info" is an indicator for whether a profile came in the second hiring task (i.e. profiles 21-30 of the 30 total profiles evaluated). The observed performance (trivia score) averages for the sample of profiles observed in Hiring Task 2 are: 38.13 (Male), 35.13 (Female), 36.95 (US), 36.53 (India), 36.84 (Under 33), 36.77 (Over 33), 36.81 (Prefer Coffee), 36.79 (Prefer Tea).

despite the minimal gap in wages and performance based on age, employers believed that young workers will significantly outperform older ones. This suggests some preference-based partiality against younger workers. Together these results highlight that a failure to account for inaccurate statistical discrimination may lead to the wrong conclusion on the source of treatment disparities.[34]

To identify whether the observed disparate treatment was driven by inaccurate statistical discrimination or animus-driven beliefs, we examined how behavior would respond to an informational intervention. Table 4 compares the differences between the

[34]In Table B6, we show that the differences in beliefs are quite similar after trimming the top and bottom five percent of the distributions of belief differences by each group. Consistent with Fig. B2, differences in beliefs about group productivities are driven by a large mass of employers with biased beliefs rather than a few employers with extreme beliefs.

two hiring rounds ("Post-Info"), the differences between wages assigned to profiles of each demographic group (e.g. "Female"), and the difference-in-differences (e.g. "Female X Post-Info"). The coefficients on "Post-Info" suggests substantial belief updating across all demographic groups, partially correcting the large level differences in the first hiring task between wages and actual group-specific productivity (a gap of roughly 5 points on average). The effect of the informational intervention on hiring decisions suggests that the majority of initial discrimination was driven by inaccurate beliefs rather than accurate statistical or preference-based sources.[35]

5 Conclusion

The study of discrimination and its motives has a rich history in economics. Separating out statistical and taste-based drivers of discrimination is a useful exercise, but as our survey of the literature illustrates, the empirical literature has thus far relied heavily on the assumption of accurate beliefs. There are many reasons to suspect that beliefs may not always be accurate. This paper formally outlines the identification problem inherent in distinguishing between belief-based and preference-based motives. A stylized experiment is used to highlight the pitfalls of not accounting for inaccurate beliefs when attempting to identify the source of discrimination, and illustrates a potential methodology for improved identification.

The results of the information intervention suggest that identifying inaccurate beliefs may have immediate policy implications for reducing discrimination. However, there are some important caveats to keep in mind when considering how this type of intervention would be implemented outside of the stylized exercise. First, such an intervention is likely feasible only in contexts where the underlying target outcome (e.g. productivity) is reliably measured and reflects the appropriate counterfactual outcome for all groups. To the first point, the accuracy of the underlying outcomes may differ by group; for example, police officers have been shown to be more likely to discount the recorded speed of

[35]There are several caveats to note when interpreting these results. Beliefs were not measured a second time. Additionally, experimenter demand may have played a role, though recent work suggests that this factor is likely small (De Quidt, Haushofer, and Roth, 2018). Finally, the change in wages could reflect an experience effect between assigning wages in the first and second hiring task. To investigate this channel, we perform a test comparing the average wages assigned in the first 10 profiles and the second 10 profiles during the initial task. We do not find evidence for an experience effect (36.86 vs. 36.72; p=.39). While we cannot fully rule out all these possible confounds, we view the information intervention as a proof of concept for the type of methodology that can be used as both an intervention for correcting beliefs and identifying belief-based discrimination from preference-based motives (e.g. animus-driven beliefs).

a white driver than a minority driver (Goncalves and Mello, 2019). To the latter point, there are contexts in which discrimination at (often unobserved) intermediate stages renders final productivity measures unreliable due to behavioral responses. For example, minority pitchers correctly anticipate discrimination by umpires and modify their behavior, resulting in a downward bias for performance measures (Parsons, Sulaeman, Yates, and Hamermesh, 2011). Studies have also documented that bias at intermediate stages can skew final productivity measures among grocery store workers (Glover, Pallais, and Pariente, 2017) and academic economists (Hengel, 2019). It is important to also take into account the underlying psychology of how people will respond to the information. Selection decisions such as hiring are rarely unidimensional. Drawing attention to a (smaller than expected) productivity gap could correct beliefs, while nonetheless increasing discrimination if it increases the salience of the gap as an input into the hiring decision. These concerns highlight the need for future tests that operationalize and examine similar informational interventions in field contexts.

Throughout the paper, we document discrimination in wages by gender (i.e. men paid more than women). Carrying out the standard outcomes-based method reveals that the gap in performance exceeds that of the gap in pay. This leads to the conclusion that there is preference-based partiality against the group that received higher wages—male workers. While taste-based discrimination is often used as a synonym for animus or prejudice against a group, this link seems misplaced when discrimination manifests as an equalizing actions (e.g. equalizing wage rates). For example, people may treat groups similarly regardless of actual or believed productivity differences due to fairness concerns. Additionally, there is often an equity-efficiency trade-off to discrimination, such that even in the absence of legal or social sanctions, an employer may wish to equalize wages across groups (for a theoretical discussion of these trade-offs in the context of racial profiling, see Durlauf (2005)). Such a concern may be especially pronounced for wages, where even abstracting away from group-level attributes, there is evidence that fairness norms may contribute to observed wage compression (e.g. Breza, Kaur, and Shamdasani (2018))

Just as decomposing the nature of belief-based discrimination has implications for policy, the same may be true for preference-based partiality. For example, if the basis for preference-based partiality is animus or prejudice, then a policy that increases contact between groups may reduce disparities (Dobbie and Fryer, 2015; Paluck, Green, and Green, 2018; Rao, 2019). By contrast, if the behavior is instead sanction- or value-

oriented, then such interventions will likely have little impact. While it is difficult to imagine a simple elicitation that would allow for a parsimonious quantitative decomposition of "tastes", survey measures may be able to make some headway in this endeavor. Such a decomposition is outside of the scope of this paper, but future work along these lines would enrich our understanding of discrimination, and help in the development of tools used to identify it and design policy.

Lastly, our findings speak to the need for continued work such as Bordalo, Coffman, Gennaioli, and Shleifer (2016) that may help to identify situations when inaccurate beliefs are especially likely to be prevalent. Two broad causes may lead to inaccurate beliefs that drive discrimination. First, research in psychology and economics has shown that heuristics and biases may generate beliefs that are systematically incorrect, leading to inaccurate stereotypes about certain groups.[36] Second, inaccurate beliefs may arise due to a lack of information—the relevant information necessary to form correct beliefs may not be available to a decision-maker. For example, an employer may have an unbiased prior belief about the productivity distributions of two groups but lack information about how selection into the job application process differs across groups, leading to inaccurate beliefs about productivity differences in the realized applicant pool. Failing to account for selection effects can also be a form of bias, as in Hübert and Little (2020) in the case of discrimination in policing. Learning will eventually mitigate inaccurate beliefs in some settings. But in other situations, there will be little or no feedback on the decisions being made, leading to learning traps in which inaccurate beliefs can persist in the long-run.[37] Further, learning may not lead to correct long-run beliefs if information is filtered through a misspecified model of the world.[38]

[36]See Schneider, Hastorf, and Ellsworth (1979); Judd and Park (1993); Hilton and Hippel (1996) and Fiske (1998) for review. Bordalo et al. (2016) present a formal model for inaccurate stereotype formation based on the representativeness heuristic. Evaluators overweight the prevalence of characteristics that differ most between groups, and end up believing that the "representative" type is more prevalent than it actually is. Biased beliefs can also arise in a dynamic learning setting when individuals use updating rules that depend on group identity (Albrecht, Von Essen, Parys, and Szech, 2013), are selective regarding which information they attend to (Schwartzstein, 2014), or have incorrect models of how others evaluate workers (Bohren et al., 2019).

[37]For example, if employers face a trade-off between learning about the productivity distribution of groups or maximizing cost-effectiveness in hiring, this can prevent full learning even though the employers are not inherently biased (Lepage, 2020).

[38]For example, confirmation bias (Rabin and Schrag, 1999), overreaction to signals (Epstein, Noor, and Sandroni, 2010), and misattribution of reference dependence (Bushong and Gagnon-Bartsch, 2022) all lead to incorrect learning in the long-run. The presence of biased agents can also lead to incorrect long-run beliefs for unbiased agents who learn from the evaluations of the biased agents but are unaware of their bias (Bohren and Hauser, 2021).

As research begins to identify situations where inaccurate beliefs are a driving factor for discrimination, future work will hopefully also begin to develop policy interventions that are able to effectively correct beliefs and reduce discrimination as a result.

References

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

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

Albrecht, K., E. Von Essen, J. Parys, and N. Szech (2013): "Updating, self-confidence, and discrimination," *European Economic Review*, 60, 144–169.

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

Anwar, S. and H. Fang (2006): "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence," *American Economic Review*, 96, 127–151.

Arnold, D., W. Dobbie, and P. Hull (2022): "Measuring racial discrimination in bail decisions," *American Economic Review*, 112, 2992–3038.

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

——— (1998): "What has economics to say about racial discrimination?" *Journal of economic perspectives*, 12, 91–100.

Ayres, I. (2002): "Outcome Tests of Racial Disparities in Police Practices," *Justice Research and Policy*, 4, 131–142.

Bar, R. and A. Zussman (2019): "Identity and Bias: Insights from Driving Tests," *Economic Journal*, 130, 1—23,.

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

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009): "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics*, 124, 1497–1540.

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

Bertrand, M. and E. Duflo (2017): "Field Experiments on Discrimination," *Handbook of Field Experiments*, 1, 110.

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.

Bohren, J. A. and D. N. Hauser (2021): "Learning With Heterogeneous Misspecified Models: Characterization and Robustness," *Econometrica*, 89, 3025–3077.

Bohren, J. A., P. Hull, and A. Imas (2022): "Systemic discrimination: Theory and measurement," .

Bohren, J. A., A. Imas, and M. Rosenberg (2019): "The Dynamics of Discrimination: Theory and Evidence," *Working Paper*.

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

Brehm, J. W. (1966): *A Theory of Psychological Reactance*, New York: Academic Press.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The morale effects of pay inequality," *Quarterly Journal of Economics*, 133, 611–663.

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

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020): "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 110, 2997–3029.

Bushong, B. and T. Gagnon-Bartsch (2022): "Learning with Misattribution of Reference Dependence," *Journal of Economic Theory*, 203, 1–45.

Cameron, S. V. and J. J. Heckman (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109, 455–499.

Charles, K. K. and J. Guryan (2011): "Studying Discrimination: Fundamental Challenges and Recent Progress," *Annual Review of Economics*, 3, 479–511.

Coffman, K. B., C. L. Exley, and M. Niederle (2021): "The role of beliefs in driving gender discrimination," *Management Science*.

De Quidt, J., J. Haushofer, and C. Roth (2018): "Measuring and bounding experimenter demand," *American Economic Review*, 108, 3266–3302.

Delavande, A. (2008): "Pill, Patch, or Shot? Subjective Expectations and Birth Control Choice," *International Economic Review*, 49, 999–1042.

Dharmapala, D. and S. L. Ross (2004): "Racial bias in motor vehicle searches: Additional theory and evidence," *Contributions to Economic Analysis and Policy*, 3, 89–111.

Dobbie, W. and R. G. Fryer (2015): "The Impact of Youth Service on Future Outcomes: Evidence from Teach for America," *The B.E. Journal of Economic Analysis and Policy*, 15, 1031–1066.

Durlauf, S. N. (2005): "Racial Profiling as a Public Policy Question: Efficiency, Equity, and Ambiguity," *American Economic Review*, 95, 132–136.

Epstein, L. G., J. Noor, and A. Sandroni (2010): "Non-Bayesian Learning," *The B.E. Journal of Theoretical Economics*, 10.

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.

Fershtman, C. and U. Gneezy (2001): "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, February, 351–377.

Fiske, S. T. (1998): "Stereotyping, prejudice, and discrimination," *The handbook of social psychology*, 2, 357–411.

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein (2018): *Channeled attention and stable errors*, Harvard Business School.

Giustinelli, P. (2016): "Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track," *International Economic Review*, 57, 573–602.

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

Gneezy, U., M. Niederle, and A. Rustichini (2003): "Performance in competitive Environments: Gender differences," *Quarterly Journal of Economics*, 1049–1074.

Goncalves, F. and S. Mello (2019): "A Few Bad Apples? Racial Bias in Policing," *American Economic Review*, 111, 1406–1441).

Grau, N. and D. Vergara (2021): "An Observational Implementation of the Outcome Test with an Application to Ethnic Prejudice in Pretrial Detentions," *Working Paper*, 1–33.

Haavelmo, T. (1944): "The Probability Approach in Econometrics," *Supplement to Econometrica*, 12, 1–115.

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

Hengel, E. (2019): "Publishing while female," *Working Paper*, 1–67.

Hilton, J. L. and W. V. Hippel (1996): "Stereotypes," *Annual Review of Psychology*, 47, 237–271.

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," Tech. rep., National Bureau of Economic Research.

Jensen, R. (2010): "The (perceived) returns to education and the demand for schooling," *Quarterly Journal of Economics*, 515–548.

Judd, C. M. and B. Park (1993): "Definition and assessment of accuracy in social stereotypes," *Psychological Review*, 100, 109–128.

Kessler, J. B., C. Low, and C. D. Sullivan (2019): "Incentivized Resume Rating: Eliciting Employer Preferences without Deception," *American Economic Review*, 109, 3173–3174.

Knowles, J., N. Persico, and P. Todd (2001): "Racial bias in motor vehicle searches: Theory and evidence," *Journal of Political Economy*, 109, 203–229.

Knox, D. E., W. Lowe, and J. Mummolo (2020): "Administrative Records Mask Racially Biased Policing," *American Political Science Review*, 1–19.

Kravitz, D. A. and J. Platania (1993): "Attitudes and Beliefs About Affirmative Action: Effects of Target and of Respondent Sex and Ethnicity," *Journal of Applied Psychology*, 78, 928–938.

Lepage, L.-P. (2020): "Endogenous Learning and the Persistence of Employer Biases in the Labor Market," Tech. rep., mimeo.

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

Manski, C. F. (2004): "Measuring Expectations," *Econometrica*, 72, 1329–1376.

Mobius, M. and T. Rosenblat (2006): "Why Beauty Matters," *American Economic Review*, 96, 222–235.

Norton, M. I. and D. Ariely (2011): "Building a better America-one wealth quintile at a time," *Perspectives on Psychological Science*, 6, 9–12.

Paluck, E. L., S. Green, and D. P. Green (2018): "The Contact Hypothesis Reevaluated," *Behavioural Public Policy*, 1–30.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011): "Strike Three: Discrimination, Incentives, and Evaluation," *American Economic Review*, 101, 1410–1435.

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

Pope, D. G. and J. R. Sydnor (2011): "What's in a Picture? Evidence of Discrimination from Prosper.com," *The Journal of Human Resources*, 46, 53–92.

Rabin, M. and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, 114, 37–82.

Rao, G. (2019): "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools," *American Economic Review*, 109, 774–809.

Schneider, D., A. Hastorf, and P. Ellsworth (1979): *Person Perception*, Reading, MA: Addison-Wesley.

Schwartzstein, J. (2014): "Selective Attention and Learning," *Journal of the European Economic Association*, 12, 1423–1452.

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.

Simoiu, C., S. Corbett-Davies, and S. Goel (2017): "The problem of infra-marginality in outcome tests for discrimination," *Annals of Applied Statistics*, 11, 1193–1216.

Tilcsik, A. (2021): "Statistical Discrimination and the Rationalization of Stereotypes," *American Sociological Review*, 86, 93–122.

Wiswall, M. and B. Zafar (2015): "Determinants of College Major Choice: Identification using an Information Experiment," *The Review of Economic Studies*, 82, 791–824.

Appendix A. Proofs from Section 3

Proof of Lemma 1. The normal distribution is the conjugate prior to a normal likelihood function. Therefore, the evaluator's posterior belief about productivity is normally distributed with mean $(\hat{\tau}_g\hat{\mu}_g + \hat{\eta}_g s)/(\hat{\tau}_g + \hat{\eta}_g)$ and variance $1/(\hat{\tau}_g + \hat{\eta}_g)$ and the optimal decision rule is $v(s, g, \theta) = 1$ iff $(\hat{\tau}_g\hat{\mu}_g + \hat{\eta}_g s)/(\hat{\tau}_g + \hat{\eta}_g) \geq u_g$. Rearranging terms yields Eq. (2). □

Proof of Lemma 2. The characterization of the set of types that exhibit equivalent discrimination follows from Eq. (2) and the discussion in the text. For the case of no discrimination, which corresponds to setting the same hiring thresholds for each group, trivially they exhibit equivalent discrimination. □

Proof of Proposition 1. Consider an evaluator of type $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g\in\{M,F\}}$ who discriminates against group F. This evaluator generates discrimination that lies on isodiscrimination curve $(s_F, s_M) = (\overline{s}(\theta, F), \overline{s}(\theta, M))$. Given that θ discriminates against group F, $s_F > s_M$. It is immediately apparent from Eq. (2) and Lemma 2 that there are a continuum of other types that exhibit equivalent discrimination. We next construct types with a single form of partiality.

Part (1): Consider a type θ' with belief neutrality, $(\hat{\mu}'_F, \hat{\tau}'_F, \hat{\eta}'_F) = (\hat{\mu}'_M, \hat{\tau}'_M, \hat{\eta}'_M)$. Let $(\hat{\mu}', \hat{\tau}', \hat{\eta}')$ denote the type's subjective beliefs for a worker from either group. Given preference parameters (u'_F, u'_M), this type hires members of group g with signals above $\overline{s}(\theta', g) = \left(\frac{\hat{\tau}'+\hat{\eta}'}{\hat{\eta}'}\right) u'_g - \frac{\hat{\tau}'}{\hat{\eta}'}\hat{\mu}'$. This type exhibits equivalent discrimination to θ if $\overline{s}(\theta', g) = s_g$ for each $g \in \{M, F\}$. Rearranging terms, this corresponds to preference parameter

$$u'_g = \left(\frac{\hat{\eta}'}{\hat{\tau}' + \hat{\eta}'}\right) s_g + \frac{\hat{\tau}'}{\hat{\tau}' + \hat{\eta}'}\hat{\mu}'$$

for group g. Note $u'_F > u'_M$ since $s_F > s_M$, so there is preference partiality against group F.

Part (2): Consider a type θ' with preference neutrality, $u'_F = u'_M$ and belief neutrality with respect to concentration and signal precision, $(\hat{\tau}'_F, \hat{\eta}'_F) = (\hat{\tau}'_M, \hat{\eta}'_M)$. Let $(u', \hat{\tau}', \hat{\eta}')$

denote these common parameters. Given subjective means $(\hat{\mu}'_M, \hat{\mu}'_F)$, this type hires members of group g with signals above $\overline{s}(\theta', g) = \left(\frac{\hat{\tau}'+\hat{\eta}'}{\hat{\eta}'}\right) u' - \frac{\hat{\tau}'}{\hat{\eta}'}\hat{\mu}'_g$. This type exhibits equivalent discrimination to θ if $\overline{s}(\theta', g) = s_g$ for each $g \in \{M, F\}$. Rearranging terms, this corresponds to subjective mean

$$\hat{\mu}'_g = \left(\frac{\hat{\tau}' + \hat{\eta}'}{\hat{\tau}'}\right) u' - \frac{\hat{\eta}'}{\hat{\tau}'} s_g$$

for group g. Note $\hat{\mu}'_F < \hat{\mu}'_M$ since $s_F > s_M$, so there is belief partiality in the form of lower expected productivity against group F.

Part (3): Consider a type θ' with preference neutrality, $u'_F = u'_M$ and belief neutrality with respect to average productivity and signal precision, $(\hat{\mu}'_F, \hat{\eta}'_F) = (\hat{\mu}'_M, \hat{\eta}'_M)$. Let $(u', \hat{\mu}', \hat{\eta}')$ denote these common parameters. Given subjective concentration of productivity $(\hat{\tau}'_M, \hat{\tau}'_F)$, this type hires members of group g with signals above $\overline{s}(\theta', g) = \left(\frac{\hat{\tau}'_g+\hat{\eta}'}{\hat{\eta}'}\right) u' - \frac{\hat{\tau}'_g}{\hat{\eta}'}\hat{\mu}'$. This type exhibits equivalent discrimination to θ if $\overline{s}(\theta', g) = s_g$ for each $g \in \{M, F\}$. Rearranging terms, this corresponds to subjective concentration

$$\hat{\tau}'_g = \hat{\eta}' \left(\frac{s_g - u'}{u' - \hat{\mu}'}\right)$$

for group g. Given $s_F > s_M$, $\hat{\eta}'(s_F - u') > \hat{\eta}'(s_M - u')$. Therefore, whether $\hat{\tau}'_F$ is greater than or less than $\hat{\tau}'_M$ depends on the sign of $u' - \hat{\mu}'$.

If $u' - \hat{\mu}' < 0$, then $\hat{\tau}'_F < \hat{\tau}'_M$ and a less concentrated subjective productivity distribution generates the discrimination against group F. The fatter low productivity tail for F relative to M means that a larger share of workers from group F fall below the threshold ex-ante. We also need to check that $\hat{\tau}'_F > 0$ for these to both be a valid precisions. This will be the case for $u' > s_F$, so that the numerator is also negative. In summary, any type with $\hat{\mu}' > s_F$, $u' \in (s_F, \hat{\mu}')$ and $\hat{\tau}'_g = \hat{\eta}' \left(\frac{s_g - u'}{u' - \hat{\mu}'}\right)$ has belief partiality in the form of lower subjective concentration for group F and exhibits equivalent discrimination to θ.

If $u' - \hat{\mu}' > 0$, then $\hat{\tau}'_F > \hat{\tau}'_M$ and a more concentrated subjective productivity distribution generates the discrimination against group F. The thinner high productivity tail for F relative to M means that a smaller share of workers from group F lie above the threshold ex-ante. We also need to check that $\hat{\tau}'_M > 0$ for these to both be valid precisions. This will be the case for $s_M > u'$, so that the numerator is also positive. In summary, any type with $\hat{\mu}' < s_M$, $u' \in (\hat{\mu}', s_M)$ and $\hat{\tau}'_g = \hat{\eta}' \left(\frac{s_g - u'}{u' - \hat{\mu}'}\right)$ has belief partiality in the form of higher subjective concentration for group F and exhibits equivalent discrimination to

θ.

Part (4): Consider a type θ' with preference neutrality, $u'_F = u'_M$ and belief neutrality with respect to average productivity and concentration, $(\hat{\mu}'_F, \hat{\tau}'_F) = (\hat{\mu}'_M, \hat{\tau}'_M)$. Let $(u', \hat{\mu}', \hat{\tau}')$ denote these common parameters. Given subjective signal precision $(\hat{\eta}'_M, \hat{\eta}'_F)$, this type hires members of group g with signals above $\overline{s}(\theta', g) = \left(\frac{\hat{\tau}'+\hat{\eta}'_g}{\hat{\eta}'_g}\right) u' - \frac{\hat{\tau}'}{\hat{\eta}'_g}\hat{\mu}'$. This type exhibits equivalent discrimination to θ if $\overline{s}(\theta', g) = s_g$ for each $g \in \{M, F\}$. Rearranging terms, this corresponds to subjective signal precision

$$\hat{\eta}'_g = \hat{\tau}' \left(\frac{u' - \hat{\mu}'}{s_g - u'}\right)$$

for group g. Given $s_F > s_M$, $s_F - u' > s_M - u'$. We need $\hat{\eta}'_g > 0$ for each g in order for these to be valid precisions. This is the case when (i) $u' - \hat{\mu}' < 0$ and $s_F - u' < 0$, which also implies $s_M - u' < 0$, or (ii) $u' - \hat{\mu}' > 0$ and $s_M - u' > 0$, which also implies $s_F - u' > 0$.

First consider case (i). In this case, $u' < \hat{\mu}'$. Further, $0 > s_F - u' > s_M - u' \Rightarrow 1/(s_M - u') > 1/(s_F - u') \Rightarrow (u' - \hat{\mu}')/(s_M - u') < (u' - \hat{\mu}')/(s_F - u')$. Therefore, $\hat{\eta}'_M < \hat{\eta}'_F$ and a higher subjective signal precision generates the discrimination against group F. In summary, any type with $\hat{\mu}' > s_F$, $u' \in (s_F, \hat{\mu}')$ and $\hat{\eta}'_g = \hat{\tau}' \left(\frac{u'-\hat{\mu}'}{s_g-u'}\right)$ has belief partiality in the form of higher subjective signal precision for group F and exhibits equivalent discrimination to θ.

Next consider case (ii). In this case, $u' > \hat{\mu}'$. Further, $s_F - u' > s_M - u' > 0 \Rightarrow 1/(s_M - u') > 1/(s_F - u') \Rightarrow (u' - \hat{\mu}')/(s_M - u') > (u' - \hat{\mu}')/(s_F - u')$. Therefore, $\hat{\eta}'_M > \hat{\eta}'_F$ and a lower subjective signal precision generates the discrimination against group F. In summary, any type with $\hat{\mu}' < s_F$, $u' \in (\hat{\mu}', s_M)$ and $\hat{\eta}'_g = \hat{\tau}' \left(\frac{u'-\hat{\mu}'}{s_g-u'}\right)$ has belief partiality in the form of lower subjective signal precision for group F and exhibits equivalent discrimination to θ. □

Lemma 3 (Identifying Type from True Distributions)**.** *Suppose a researcher identifies the hiring rules (s_M, s_F) and true distributions (μ_g, τ_g, η_g) for $g \in \{M, F\}$. Assume an evaluator has accurate beliefs. Then the evaluator's preference parameter is identified as:*

$$u_g = \left(\frac{\eta_g}{\tau_g + \eta_g}\right) s_g + \left(\frac{\tau_g}{\tau_g + \eta_g}\right) \mu_g. \tag{6}$$

and the evaluator's beliefs are identified as $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) = (\mu_g, \tau_g, \eta_g)$ for $g \in \{M, F\}$.

Proof of Lemma 3. Suppose the evaluator has type $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g\in\{M,F\}}$. This evaluator exhibits discrimination that lies on isodiscrimination curve $(s_F, s_M) = (\overline{s}(\theta, F), \overline{s}(\theta, M))$. Suppose the researcher identifies the hiring rules (s_F, s_M) and the true productivity and signal distributions (μ_g, τ_g, η_g) for each group g. Under the assumption of accurate beliefs, i.e. $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) = (\mu_g, \tau_g, \eta_g)$, solving Eq. (2) for u_g uniquely identifies the preference parameters as Eq. (6). Therefore, the evaluator's type is identified. □

Proof of Proposition 2. Given true productivity and signal distributions $(\mu_g, \tau_g, \eta_g)_{g\in\{M,F\}}$, suppose the evaluator has type $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g\in\{M,F\}}$ with inaccurate beliefs, $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g) \neq (\mu_g, \tau_g, \eta_g)$. This evaluator exhibits discrimination that lies on isodiscrimination curve $(s_F, s_M) = (\overline{s}(\theta, F), \overline{s}(\theta, M))$. Suppose a researcher identifies the hiring rules (s_F, s_M) and the true productivity and signal distributions (μ_g, τ_g, η_g) for each group g. When the researcher assumes belief are accurate, i.e. the evaluator is a type θ' with beliefs $(\hat{\mu}'_g, \hat{\tau}'_g, \hat{\eta}'_g) = (\mu_g, \tau_g, \eta_g)$, then from Lemma 3, the researcher concludes that the evaluator has preference parameter

$$u'_g = \left(\frac{\eta_g}{\tau_g + \eta_g}\right) s_g + \left(\frac{\tau_g}{\tau_g + \eta_g}\right) \mu_g. \tag{7}$$

In contrast, the the true preference parameter satisfies

$$u_g = \left(\frac{\hat{\eta}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) s_g + \left(\frac{\hat{\tau}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) \hat{\mu}_g. \tag{8}$$

When beliefs are inaccurate, this identified preference parameter is equal to the true parameter, $u'_g = u_g$, if and only if

$$\mu_g = \left(\frac{\tau_g + \eta_g}{\tau_g}\right) \left[\left(\frac{\hat{\eta}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) s_g - \left(\frac{\eta_g}{\tau_g + \eta_g}\right) s_g + \left(\frac{\hat{\tau}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) \hat{\mu}_g\right]. \tag{9}$$

Therefore, the preference parameter is misidentified for a generic set of true beliefs (μ_g, τ_g, η_g) and evaluator types $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g\in\{M,F\}}$, $u'_g \neq u_g$.

Let $\theta^* = (u_g, \mu_g, \tau_g, \eta_g)_{g\in\{M,F\}}$ denote the type with accurate beliefs and the same preferences as θ. Suppose type θ's inaccurate beliefs increase discrimination against group F, i.e. $\overline{s}(\theta, F) \geq \overline{s}(\theta^*, F)$ and $\overline{s}(\theta^*, M) \geq \overline{s}(\theta, M)$ with at least one strict inequality. Then given the observed hiring rules are consistent with type θ, i.e. $s_F = \overline{s}(\theta, F)$, $s_F \geq \overline{s}(\theta^*, F) = \frac{\tau_F + \eta_F}{\eta_F} u_F - \frac{\tau_F}{\eta_F} \mu_F$. Combining this inequality with Eq. (7) establishes

that

$$u'_F = \left(\frac{\eta_F}{\tau_F + \eta_F}\right) s_F + \left(\frac{\tau_F}{\tau_F + \eta_F}\right) \mu_F \geq u_F. \tag{10}$$

Similarly, $u'_M \leq u_M$, with a strict inequality for at least one of the expressions. Therefore, the researcher overestimates the preference parameter for group F and/or underestimates the preference parameter for group M, leading her to overestimate the preference partiality against group F. The proof for the case of decreasing discrimination is analogous. □

Proof of Proposition 3. Suppose the researcher identifies the hiring rules (s_F, s_M) and the true productivity and signal distributions (μ_g, τ_g, η_g) for each group g. From Eq. (2), for any $u \in \mathbb{R}$, the corresponding accurate statistical discriminator with preferences $u_M = u_F = u$ lies on isodiscrimination curve (s'_F, s'_M) with $s'_g = \left(\frac{\tau_g + \eta_g}{\eta_g}\right) u - \frac{\tau_g}{\eta_g}\mu_g$. If $\frac{\tau_M \mu_M + \eta_M s_M}{\tau_M + \eta_M} \neq \frac{\tau_F \mu_F + \eta_F s_F}{\tau_F + \eta_F}$, then there is no u such that $(s'_F, s'_M) = (s_F, s_M)$, i.e. there is no u such that an accurate statistical discriminator with preference parameter u exhibits discrimination that is consistent with the observed hiring rules. □

Proof of Proposition 4. Suppose the researcher identifies the hiring rules (s_F, s_M) and the subjective productivity and signal distributions $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)$ for each group g. Solving Eq. (2) for u_g uniquely identifies the preference parameters (u_F, u_M) as

$$u_g = \left(\frac{\hat{\eta}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) s_g + \left(\frac{\hat{\tau}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) \hat{\mu}_g. \tag{11}$$

Therefore, the evaluator's type is identified. □

Proof of Proposition 5. Given a signal with precision $\eta > 0$, observing $x \geq 1$ draws of the signal is equivalent to observing a single signal that is normally distributed with precision $x\eta$. Suppose the researcher identifies the hiring rules $(s_{F,1}, s_{M,1})$ when she observes x_1 signal draws, and hiring rules $(s_{F,2}, s_{M,2})$ when she observes $x_2 \neq x_1$ signal draws. Then from Eq. (2), this is consistent with any type $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g \in \{M,F\}}$ such that

$$\frac{\hat{\tau}_g + x_i \hat{\eta}_g}{x_i \hat{\eta}_g} u_g - \frac{\hat{\tau}_g}{x_i \hat{\eta}_g} \hat{\mu}_g = s_{g,i}. \tag{12}$$

for $i = 1, 2$ and $g \in \{M, F\}$. Rearranging terms,

$$\left(\frac{\hat{\tau}_g}{\hat{\eta}_g} + x_i\right) u_g = x_i s_{g,i} + \frac{\hat{\tau}_g}{\hat{\eta}_g}\hat{\mu}_g. \tag{13}$$

Subtracting Eq. (13) evaluated at x_2 from Eq. (13) evaluated at x_1 and solving for u_g identifies the evaluator's preferences as

$$u_g = \frac{x_1 s_{g,1} - x_2 s_{g,2}}{x_1 - x_2}. \tag{14}$$

However, multiple sets of beliefs $(\hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g\in\{M,F\}}$ can be consistent with these hiring rules. To see this, suppose type $\theta = (u_g, \hat{\mu}_g, \hat{\tau}_g, \hat{\eta}_g)_{g\in\{M,F\}}$ is consistent with the observed hiring rule. This implies u_g satisfies Eq. (14). Then $\theta' = (u_g, \hat{\mu}'_g, \hat{\tau}'_g, \hat{\eta}'_g)_{g\in\{M,F\}}$ exhibits equivalent discrimination to θ when observing x signal draws if

$$\frac{\hat{\tau}_g + x\hat{\eta}_g}{x\hat{\eta}_g} u_g - \frac{\hat{\tau}_g}{x\hat{\eta}_g}\hat{\mu}_g = \frac{\hat{\tau}'_g + x\hat{\eta}'_g}{x\hat{\eta}'_g} u_g - \frac{\hat{\tau}'_g}{x\hat{\eta}'_g}\hat{\mu}'_g \tag{15}$$

for $g \in \{M, F\}$. Rearranging terms, this is equivalent to

$$\frac{\hat{\tau}_g(u_g - \hat{\mu}_g)}{\hat{\eta}_g} = \frac{\hat{\tau}'_g(u_g - \hat{\mu}'_g)}{\hat{\eta}'_g} \tag{16}$$

which is independent of x. It is readily apparent that a continuum of types $\theta' = (u_g, \hat{\mu}'_g, \hat{\tau}'_g, \hat{\eta}'_g)_{g\in\{M,F\}}$ can satisfy this equation. For example, types with $\hat{\mu}'_g = \hat{\mu}_g$ and that preserve the ratio of the precisions, $\hat{\tau}'_g/\hat{\eta}'_g = \hat{\tau}_g/\hat{\eta}_g$, exhibit equivalent discrimination to θ. Therefore, the evaluator's beliefs are not identified.

Moreover, since Eq. (16) is independent of x, any $\theta' = (u_g, \hat{\mu}'_g, \hat{\tau}'_g, \hat{\eta}'_g)_{g\in\{M,F\}}$ that satisfies Eq. (15) also exhibits equivalent discrimination to θ for any other informational treatment $x \notin \{x_1, x_2\}$. In other words, additional informational treatments will provide no additional scope to identify beliefs.

□

Appendix B. Additional Tables and Figures

Table B1. Summary Statistics

	Total	**Male**	**Female**	**US**	**India**	**Under 33**	**Over 33**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Worker							
Trivia Score	36.95	38.32	35.28	37.14	36.58	37.10	36.79
	(8.73)	(8.52)	(8.70)	(8.93)	(8.31)	(8.55)	(8.94)
Survey Duration (Minutes)	18.82	19.03	18.56	16.19	24.04	20.25	17.18
	(10.39)	(10.52)	(10.25)	(8.12)	(12.31)	(11.82)	(8.20)
Prefer Tea (Yes=1)	0.39	0.38	0.41	0.37	0.44	0.42	0.36
	(0.49)	(0.49)	(0.49)	(0.48)	(0.50)	(0.49)	(0.48)
Age (Worker)	35.89	35.30	36.62	38.55	30.61	27.38	45.61
	(11.57)	(11.27)	(11.91)	(12.16)	(8.01)	(3.50)	(9.76)
Female (Yes=1)	0.45	0.00	1.00	0.52	0.32	0.43	0.48
	(0.50)	(0.00)	(0.00)	(0.50)	(0.47)	(0.50)	(0.50)
From India (Yes=1)	0.33	0.42	0.23	0.00	1.00	0.47	0.18
	(0.47)	(0.49)	(0.42)	(0.00)	(0.00)	(0.50)	(0.39)
# Observations	589	324	265	392	197	314	275
Panel B: Employer							
Survey Duration (Minutes)	23.09	23.59	22.37	19.08	31.60	22.53	23.87
	(17.23)	(15.57)	(19.43)	(11.70)	(23.04)	(19.00)	(14.44)
College Education or Above	0.67	0.70	0.62	0.56	0.90	0.67	0.67
	(0.47)	(0.46)	(0.49)	(0.50)	(0.30)	(0.47)	(0.47)
Age (Employer)	34.36	32.66	36.88	35.73	31.46	27.09	44.36
	(11.02)	(9.92)	(12.07)	(11.63)	(8.96)	(3.59)	(9.91)
Female (Yes=1)	0.40	0.00	1.00	0.49	0.23	0.34	0.49
	(0.49)	(0.00)	(0.00)	(0.50)	(0.42)	(0.47)	(0.50)
From India (Yes=1)	0.32	0.41	0.19	0.00	1.00	0.40	0.29
	(0.47)	(0.49)	(0.39)	(0.00)	(0.00)	(0.49)	(0.41)
# Observations	577	344	233	392	185	334	243

Notes: Standard deviations in parentheses. One observation per worker (survey 1) or employer (survey 2).

Figure B1. Kernel Densities of Productivities (Trivia Scores) by Group

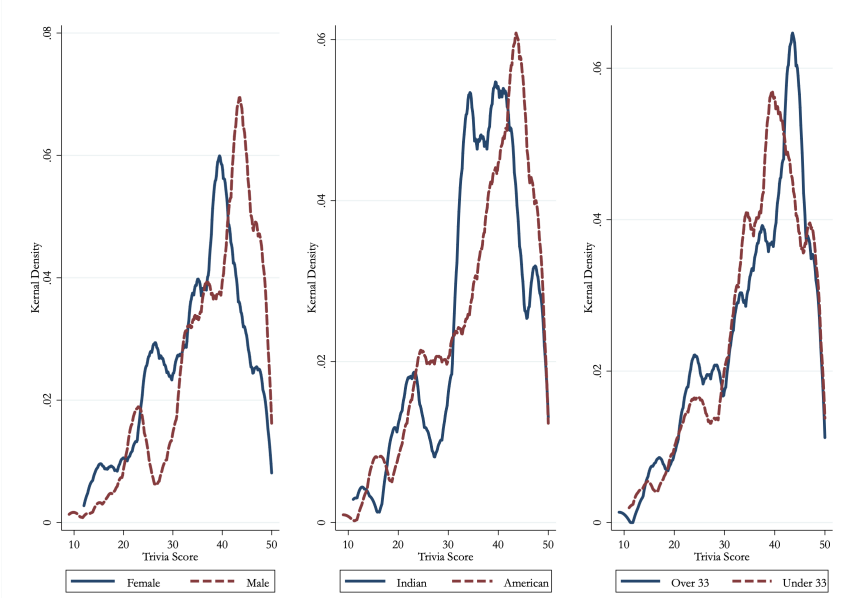


Figure B2. Kernel Densities of Beliefs about Differences by Group (Within-Employer)

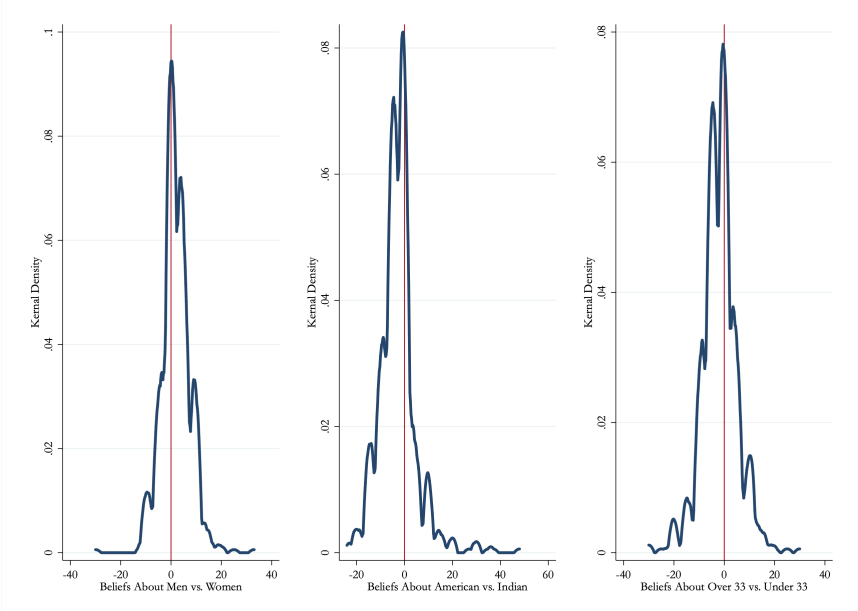


Table B2. Discrimination in Wages, by Employee Characteristics (Hiring Task 1)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-1.05***			-0.66*	-0.78**	-0.80**	-0.62**	-0.48*
	(0.38)			(0.37)	(0.33)	(0.33)	(0.27)	(0.27)
Indian		2.14***		2.01***	2.03***	2.00***	1.09***	1.25***
		(0.41)		(0.43)	(0.38)	(0.38)	(0.32)	(0.32)
Over 33			-0.54	0.06	0.31	0.32	0.35	0.35
			(0.39)	(0.39)	(0.35)	(0.35)	(0.29)	(0.28)
Prefers Tea						0.37		0.37
						(0.32)		(0.26)
Fav Subject: Math							5.31***	5.24***
							(0.37)	(0.37)
Fav Color: Blue								0.18
								(0.28)
Fav Sport: Football								0.76**
								(0.30)
Fav Movie: Popular								1.05***
								(0.31)
N	11,540	11,540	11,540	11,540	11,540	11,540	11,540	11,540
R^2	0.00	0.01	0.00	0.01	0.49	0.49	0.52	0.52
DepVarMean	31.90	30.71	31.67	30.71	30.71	30.71	30.71	30.71
Employer FE?	No	No	No	No	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses, two-way clustered by employer and worker. "DepVarMean" is the mean of the dependent variable (wage WTP) in the omitted group (e.g. Male Workers for column (1)). To control for the non-target characteristics of profiles, we needed to turn the free responses into numeric variables – we chose to do so by binarizing each (other than Coffee/Tea preference, which was already binary). For "Favorite High School Subject" we defined this variable as equal to 1 if the worker mentioned "math" (e.g. "maths" or "MATHEMATICS") in their response (25.8% of workers). For color, we used the most common response, those containing "blue" (38.9%), for sport we used the most common sport of "football" or "soccer" (26.8%) and for favorite movie we included any movie that was mentioned by at least 5 workers (17.0%, i.e. movies containing the words "titanic", "star wars", "shawshank", "avatar", "inception", "rings", "matrix", or "princess bride")

Table B3. In-Group Bias Test (Hiring Task 1)

	(1)	(2)	(3)	(4)
Female Worker	-1.42***			-1.20***
	(0.37)			(0.37)
Female Employer	1.78**			1.91***
	(0.69)			(0.72)
Female Worker X Employer	0.26			0.41
	(0.44)			(0.44)
Indian Worker		2.04***		1.88***
		(0.44)		(0.45)
Indian Employer		0.99		1.70**
		(0.71)		(0.75)
Indian Worker X Employer		-0.79		-0.82
		(0.51)		(0.51)
Over 33 Worker			-0.86**	-0.39
			(0.37)	(0.37)
Over 33 Employer			0.31	0.22
			(0.69)	(0.71)
Over 33 Worker X Employer			1.10***	1.19***
			(0.40)	(0.40)
N	17,310	17,310	17,310	17,310
R^2	0.01	0.01	0.00	0.02
DepVarMean	31.90	30.71	31.67	31.67

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors in parentheses, two-way clustered by employer and worker. "DepVarMean" is the mean of the dependent variable (wage WTP) in the omitted group (e.g. Male Workers evaluated by Male Employers for column (1)).

Table B4. In-Group vs. Out-Group Beliefs about Productivity by Employee Characteristics

	Out Group	**In Group**	**Diff.**	**p-val**	**#Obs. Out**	**#Obs. In**
	(1)	(2)	(3)	(4)	(5)	(6)
Prediction for Female Workers	31.70	32.79	-1.09	0.13	344	233
	(8.78)	(7.81)				
Prediction for Male Workers	34.68	33.60	1.09	0.12	233	344
	(6.59)	(9.20)				
Prediction for Indian Workers	36.09	32.06	4.04	0.00	392	185
	(7.10)	(12.67)				
Prediction for US Workers	30.46	32.84	-2.38	0.00	185	392
	(12.04)	(6.15)				
Prediction for Over 33 Workers	30.92	32.47	-1.55	0.04	334	243
	(9.82)	(7.66)				
Prediction for Under 33 Workers	33.85	33.09	0.77	0.31	243	334
	(7.03)	(10.14)				

Notes: Standard deviations in parentheses. "In-Group" refers to a match in the characteristic between the employer and the group of workers over which they are making a prediction, e.g. column 2, row 1 is the average prediction made by female employers about the average productivity of female workers.

Table B6. Beliefs about Productivity by Employee Characteristics, Trimmed

	(1) Group 1	(2) (1 or 2) 2	(3) Diff. [(1)-(2)]	(4) P-Val
Gender (1 = Male , 2 = Female)	34.26	32.30	1.96	0.00
	(8.23)	(8.20)		
Country (1 = US , 2 = India)	32.00	35.21	-3.22	0.00
	(8.49)	(8.85)		
Age (1 = Under 33 , 2 = Over 33)	33.42	31.78	1.64	0.00
	(8.84)	(8.83)		

Notes: This table repeats Table 3 after trimming the top and bottom 5 percent of observations by the within-employer difference in beliefs about the two groups (e.g. on the Male - Female difference for the first row). Standard deviations in parentheses. One observation per employer combination. Column (4) shows the p-value from regression of the outcome on a dummy variable for group membership, with standard errors two-way clustered by employer and worker. # Observations = 528 (Gender), 541 (Country), and 528 (Age).

Table B5. Effects of Large Incentives for Accurate Predictions

	Incentivized?		**Diff.**	**p-val**
	No	*Yes*		
	(1)	(2)	(3)	(4)
Prediction for Female Workers	32.36	31.93	0.44	0.53
	(7.71)	(9.08)		
Prediction for Male Workers	34.22	33.86	0.36	0.60
	(7.37)	(9.08)		
Prediction for Indian Workers	35.29	34.31	0.98	0.21
	(8.49)	(10.30)		
Prediction for US Workers	32.28	31.87	0.41	0.56
	(8.21)	(8.90)		
Prediction for Over 33 Workers	31.95	31.19	0.75	0.32
	(8.39)	(9.58)		
Prediction for Under 33 Workers	33.73	33.09	0.64	0.39
	(8.58)	(9.35)		
# Observations	290	287		

Notes: Standard deviations in parentheses. One observation per employer. The joint f-statistic from regression of an indicator for the "Incentivized" treatment on set of employer observable characteristics in Table B1, Panel B (duration, education, age, female, from India) is 1.25 (p=0.286).

Inaccurate Statistical Discrimination: An Identification Problem

J. Aislinn Bohren, Kareem Haggag, Alex Imas, and Devin G. Pope.

December 16, 2022

Supplemental Material: Literature Survey

This file further discusses empirical work from the literature survey in Section 2 that considers inaccurate beliefs and then lists the citation for each paper included in the survey.

Methodology of Literature Survey

We classified papers as "discuss taste-based versus statistical source" if preference versus belief-based motives for the documented discrimination were discussed in the text, and as "test for taste-based versus statistical source" if the paper either explicitly tested between different models of preference versus belief-based discrimination or implicitly tested the predictions of a belief-based model while taking the taste-based model as the null hypothesis. If a paper mentioned inaccurate or biased beliefs as a potential source of discrimination, it was classified as "discuss accurate versus inaccurate beliefs." Papers that tested whether inaccurate beliefs could be driving discrimination, either by directly eliciting beliefs or through other tests, were classified as "test for inaccurate beliefs." Finally, papers that elicited beliefs were classified as "measure beliefs." Three of the seven papers in this category did not test whether these elicited beliefs were accurate.

Of the papers that consider inaccurate beliefs in identification, List (2004); Hedegaard and Tyran (2018); Mobius and Rosenblat (2006) measure beliefs either directly or indirectly. List (2004) studies discrimination in bargaining and negotiations in the context of a sports card market. Dealer perceptions of buyers' reservation prices (RPs) are assessed by presenting them with actual RP distributions and asking them match the distributions to buyer sub-groups. The paper argues that observed disparities in bargaining outcomes are due to statistical discrimination because the dealers' matching rates are significantly higher than chance, with higher accuracy for more experienced dealers. Mobius and Rosenblat (2006) investigate the beauty premium in a laboratory experiment. Workers are hired by employers to solve maze puzzles. Despite no productivity differences on the task based on attractiveness, the authors document a significant beauty premium. Eliciting beliefs shows that both visual and oral interactions lead employers to form mistaken perceptions that attractive workers are more productive. Hedegaard

and Tyran (2018) study preferences for co-workers as a function of their group identify and productivity. They find that people have a significant preference for working with a member of the same ethnicity. To provide evidence that this is due to taste-based discrimination, the authors elicit productivity beliefs from a separate group of subjects and show that beliefs are qualitatively accurate, and thus cannot explain the observed differential treatment. Agan and Starr (2017); Arnold et al. (2018) derive predictions from a specific structural model of biased beliefs and takes these predictions to the data. Agan and Starr (2017) run a correspondence study to examine how ban-the-box policies affect call back rates for minority applicants. They use a model to estimate employer priors of criminality by group identity and compare those estimates to actual criminality estimates found in the literature. The discrepancy between those statistics is used to argue that employers have incorrect stereotypes. Arnold et al. (2018) examine racial bias in judicial decisions by comparing release tendencies and pretrial misconduct rates as a function of group identity. Comparing pretrial misconduct rates of the marginal defendant suggests racial bias. To explore the source of this bias, the authors estimate the misconduct risk distributions by group identity, arguing that if judges are subject to the representativeness heuristic as in Bordalo et al. (2016), then bias against Black defendents are likely due to stereotypes. Finally, Fershtman and Gneezy (2001) use a laboratory experiment to study discrimination in Israel. Behavior in the trust game—where payment is based on the actions of one's partner—versus a dictator game—where payment is strictly a function of one player—is used to study the source of discrimination. Differential treatment is observed in the former game but not the latter, which is used to argue that discrimination is due to mistaken stereotypes rather than animus.[39]

Method. We now proceed to outline the method that we used to determine which papers to include in the survey and the data that we collected for each paper.

Inclusion Criteria. We focused on empirical papers published between 1990 and 2018 in the following journals: American Economic Journal: Applied, American Economic Journal: Policy, American Economic Review (excluding the Papers & Proceedings issue), Econometrica, Journal of the European Economic Association, Journal of Labor Economics, Journal of Political Economy, the Quarterly Journal of Economics, Review

[39]Beyond these seven papers, we believe that inaccurate statistical discrimination is a plausible, untested interpretation in the majority of the remaining studies. One possible exception is Hjort (2014), in which a likely shock to preferences (ethnic conflict following an election) leads to an increase in discrimination, which is interpreted as evidence for taste-based discrimination. However, this does not preclude the possibility of inaccurate statistical discrimination as an additional driver of discrimination both before and after the preference shock.

Table B7. Publications by Journal and Decade

	Number of Papers			
	1990-99	**2000-09**	**2010-2018**	**Total**
AEJ: Applied	0	1	7	8
AEJ: Policy	0	0	2	2
AER	4	7	6	17
EMA	0	0	0	0
JEEA	0	1	1	2
JLE	2	8	12	22
JPE	2	6	1	9
ReStud	1	2	3	6
ReStat	5	6	11	22
QJE	4	4	9	17
Total	**18**	**35**	**52**	**105**

of Economic Studies, and Review of Economics and Statistics. We acknowledge that the economics literature on discrimination includes important contributions from other journals. We restricted attention to these ten journals as a representative sample in order for the scope of the survey to include a manageable number of papers.

We proceeded in two steps to determine whether to include a paper published in the relevant time frame and journals. First, in each journal, we searched for all empirical papers that had at least one of the search terms *{discrimination, prejudice, bias, biases, biased, disparity, disparities, stereotype, stereotypes, premium}* in the title, or at least one of the search terms *{discrimination, prejudice}* in the abstract, or at least one of the search terms from *{racial, race, gender, sex, ethnic, religious, beauty}* and *{bias, biased, disparity, stereotype, stereotypes, premium}* in the abstract. Second, we restricted attention to papers that attempted to causally document differential treatment of individuals based on their group identity. This eliminated papers on unrelated topics, including the industrial organization literature on *price* discrimination, the financial literature on the *risk* premium, theoretical models, and the experimental literature that documents behavioral differences such as gender differences in risk preferences.[40]

[40]We also excluded some papers that met our objective criteria but which we viewed as not relevant to the spirit of the exercise. More specifically, we excluded papers that could not be classified as either a "Yes" or "No" for the criteria outlined in Table 1. For example, Gneezy, Niederle, and Rustichini (2003) examine behavioral differences between men and women but do not study discrimination per se. Similarly, Cameron and Heckman (2001) examine the extent to which the racial and ethnic gap in college attendance can be explained by long-run versus short-run factors but do not address discrimination.

Data Collection. For each paper that met our inclusion criteria, we recorded the following information: data source (laboratory experiment, field experiment, audit or correspondence study, observational data study, other), empirical method (reduced form analysis, structural analysis), group identity of interest (race, gender, ethnicity, religion, sexuality, class/income, other), domain of study (labor market, legal, education, financial, consumer purchases—non-financial, evaluations, other), measure of discrimination (i.e. difference in call back rates), whether the paper distinguishes between taste-based and statistical discrimination, whether the paper distinguishes between accurate and inaccurate statistical discrimination, whether discrimination was documented, whether the study identified the source of discrimination, and whether the study measured beliefs about an individual's predicted attribute by group identity.

Summary Statistics. We found 105 papers that met our inclusion criteria. Table B7 lists the number of papers broken down by journal and decade of publication. The full list of papers is included in the Supplemental Material. Out of the papers surveyed, 11 conducted audit or correspondence studies, 7 conducted another type of field experiment, 3 conducted a laboratory experiment and 84 analyzed observational data.

Discrimination was studied for a variety of group identities and in a variety of domains. The most frequent group identities were race (58 papers) and gender (37 papers), followed by physical traits / appearance (7 papers) and ethnicity (6 papers). The most frequent domain was labor markets (58 papers), followed by legal contexts (12 papers), education (9 papers), non-financial consumer markets (6 papers) and financial markets (5 papers). Table B8 summarizes the papers by group identity and domain. Some papers in the survey studied multiple group identities or domains; therefore, some papers are counted in multiple rows of the table.

Table B8. Type and Domain of Discrimination

	All Papers	**Evidence of Discrimination**	
	# Papers	*# Papers*	*% Total*
Group Identity			
Race	58	56	96.6%
Gender	37	35	94.6%
Ethnicity	6	6	100.0%
Religion	1	1	100.0%
Sexuality	1	1	100.0%
Class/Income	1	1	100.0%
Physical Traits / Appearance	7	7	100.0%
Other	5	5	100.0%
Domain of Discrimination			
Labor Market	58	57	98.3%
Legal	12	12	100.0%
Education	9	9	100.0%
Financial	5	4	80.0%
Consumer Markets (not financial)	6	6	100.0%
Other	17	16	94.1%

Citations of Papers Included in Literature Survey

Abrevaya, J. and D. S. Hamermesh (2012): "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?" *The Review of Economics and Statistics*, 94, 202–207.

Acemoglu, D. and J. Angrist (2001): "Consequences of Employment Protection? The Case of the Americans with Disabilities Act," *Journal of Political Economy*, 109, 915–957.

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

Alan, S., S. Ertac, and I. Mumcu (2018): "Gender Stereotypes in the Classroom and Effects on Achievement," *The Review of Economics and Statistics*, 100, 876–890.

Alesina, A. and E. L. Ferrara (2014): "A Test of Racial Bias in Capital Sentencing," *The American Economic Review*, 104, 3397–3433.

Altonji, J. G. and C. R. Pierret (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

Antecol, H. and P. Kuhn (2000): "Gender as an Impediment to Labor Market Success: Why Do Young Women Report Greater Harm?" *Journal of Labor Economics*, 18, 702–728.

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

Anwar, S., P. Bayer, and R. Hjalmarsson (2012): "The Impact of Jury Race in Criminal Trials," *The Quarterly Journal of Economics*, 127, 1017–1055.

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

Arai, M. and P. Skogman Thoursie (2009): "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings," *Journal of Labor Economics*, 27, 127–147.

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

Aslund, O., L. Hensvik, and O. N. Skans (2014): "Seeking Similarity: How Immigrants and Natives Manage in the Labor Market," *Journal of Labor Economics*, 32, 405–441.

Ayres, I. and P. Siegelman (1995): "Race and Gender Discrimination in Bargaining for a New Car," *The American Economic Review*, 85, 304–321.

Bagues, M. F. and B. Esteve-Volart (2010): "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *The Review of Economic Studies*, 77, 1301–1328.

Baldwin, M. and W. G. Johnson (1992): "Estimating the Employment Effects of Wage Discrimination," *The Review of Economics and Statistics*, 74, 446–455.

Bar, R. and A. Zussman (2017): "Customer Discrimination: Evidence from Israel," *Journal of Labor Economics*, 35, 1031–1059.

Barcellos, S. H., L. S. Carvalho, and A. Lleras-Muney (2014): "Child Gender and Parental Investments In India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics*, 6, 157–189.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *The American Economic Review*, 106, 1437–1475.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009): "Powerful Women: Does Exposure Reduce Bias?" *The Quarterly Journal of Economics*, 124, 1497–1540.

Beaman, L., N. Keleher, and J. Magruder (2018): "Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi," *Journal of Labor Economics*, 36, 121–157.

Berkovec, J. A., G. B. Canner, S. A. Gabriel, and T. H. Hannan (1998): "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending," *The Review of Economics and Statistics*, 80, 241–250.

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

Biddle, J. E. and D. S. Hamermesh (1998): "Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics*, 16, 172–201.

Black, S. E. and P. E. Strahan (2001): "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry," *The American Economic Review*, 91, 814–831.

Blanchflower, D. G., P. B. Levine, and D. J. Zimmerman (2003): "Discrimination in the Small-Business Credit Market," *The Review of Economics and Statistics*, 85, 930–943.

Bollinger, C. R. (2003): "Measurement Error in Human Capital and the Black-White Wage Gap," *The Review of Economics and Statistics*, 85, 578–585.

Botelho, F., R. A. Madeira, and M. A. Rangel (2015): "Racial Discrimination in Grading: Evidence from Brazil," *American Economic Journal: Applied Economics*, 7, 37–52.

Breda, T. and S. T. Ly (2015): "Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France," *American Economic Journal: Applied Economics*, 7, 53–75.

Brevoort, K. P. (2011): "Credit Card Redlining Revisited," *The Review of Economics and Statistics*, 93, 714–724.

Brueckner, J. and Y. Zenou (2003): "Space and Unemployment: The Labor-Market Effects of Spatial Mismatch," *Journal of Labor Economics*, 21, 242–262.

Burgess, S. and E. Greaves (2013): "Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities," *Journal of Labor Economics*, 31, 535–576.

Butcher, K. F., K. H. Park, and A. M. Piehl (2017): "Comparing Apples to Oranges: Differences in Women's and Men's Incarceration and Sentencing Outcomes," *Journal of Labor Economics*, 35, S201–S234.

Carruthers, C. K. and M. H. Wanamaker (2017): "Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap," *Journal of Labor Economics*, 35, 655–696.

Case, A. and C. Paxson (2008): "Stature and Status: Height, Ability, and Labor Market Outcomes," *Journal of Political Economy*, 116, 499–532.

Charles, K. and J. Guryan (2008): "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination," *Journal of Political Economy*, 116, 773–809.

Combes, P.-P., B. Decreuse, M. Laouénan, and A. Trannoy (2016): "Customer Discrimination and Employment Outcomes: Theory and Evidence from the French Labor Market," *Journal of Labor Economics*, 34, 107–160.

Dahl, G. B. and E. Moretti (2008): "The Demand for Sons," *The Review of Economic Studies*, 75, 1085–1120.

Donald, S. G. and D. S. Hamermesh (2006): "What Is Discrimination? Gender in the American Economic Association, 1935-2004," *The American Economic Review*, 96, 1283–1292.

Eckstein, Z. and K. I. Wolpin (1999): "Estimating the Effect of Racial Discrimination on First Job Wage Offers," *The Review of Economics and Statistics*, 81, 384–392.

Edelman, B., M. Luca, and D. Svirsky (2017): "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment," *American Economic Journal: Applied Economics*, 9, 1–22.

Elliehausen, G. E. and E. C. Lawrence (1990): "Discrimination in Consumer Lending," *The Review of Economics and Statistics*, 72, 156–160.

Ewens, M., B. Tomlin, and L. C. Wang (2014): "Statistical Discrimination or Prejudice? A Large Sample Field Experiment," *The Review of Economics and Statistics*, 96, 119–134.

Fershtman, C. and U. Gneezy (2001): "Discrimination in a Segmented Society: An Experimental Approach," *The Quarterly Journal of Economics*, 116, 351–377.

Foote, C., W. Whatley, and G. Wright (2003): "Arbitraging a Discriminatory Labor Market: Black Workers at the Ford Motor Company, 1918–1947," *Journal of Labor Economics*, 21, 493–532.

Foster, A. D. and M. R. Rosenzweig (1996): "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market," *The Review of Economic Studies*, 63, 347–374.

Fryer, R. G. and S. D. Levitt (2010): "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, 2, 210–240.

Gardeazabal, J. and A. Ugidos (2004): "More on Identification in Detailed Wage Decompositions," *The Review of Economics and Statistics*, 86, 1034–1036.

Gayle, G.-L. and L. Golan (2012): "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968-1997," *The Review of Economic Studies*, 79, 227–267.

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.

Goldin, C. and C. Rouse (2000): "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *The American Economic Review*, 90, 715–741.

Goldsmith, A. H., D. Hamilton, and W. Darity (2006): "Shades of Discrimination: Skin Tone and Wages," *The American Economic Review*, 96, 242–245.

Gong, J., Y. Lu, and H. Song (2018): "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes," *Journal of Labor Economics*, 36, 743–778.

Hamermesh, D. S. and J. E. Biddle (1994): "Beauty and the Labor Market," *The American Economic Review*, 84, 1174–1194.

Hanna, R. N. and L. L. Linden (2012): "Discrimination in Grading," *American Economic Journal: Economic Policy*, 4, 146–168.

Heap, S. P. H. and D. J. Zizzo (2009): "The Value of Groups," *The American Economic Review*, 99, 295–323.

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

Hendricks, W., L. DeBrock, and R. Koenker (2003): "Uncertainty, Hiring, and Subsequent Performance: The NFL Draft," *Journal of Labor Economics*, 21, 857–886.

Hersch, J. (2008): "Profiling the New Immigrant Worker: The Effects of Skin Color and Height," *Journal of Labor Economics*, 26, 345–386.

Heywood, J. S. and D. Parent (2012): "Performance Pay and the White-Black Wage Gap," *Journal of Labor Economics*, 30, 249–290.

Hirsch, B. and D. Macpherson (2004): "Wages, Sorting on Skill, and the Racial Composition of Jobs," *Journal of Labor Economics*, 22, 189–210.

Hirsch, B., T. Schank, and C. Schnabel (2010): "Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany," *Journal of Labor Economics*, 28, 291–330.

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

Holzer, H. J. and K. R. Ihlanfeldt (1998): "Customer Discrimination and Employment Outcomes for Minority Workers," *The Quarterly Journal of Economics*, 113, 835–867.

Ichino, A. and E. Moretti (2009): "Biological Gender Differences, Absenteeism, and the Earnings Gap," *American Economic Journal: Applied Economics*, 1, 183–218.

Ihlanfeldt, K. R. and M. V. Young (1994): "Intrametropolitan Variation in Wage Rates: The Case of Atlanta Fast-Food Restaurant Workers," *The Review of Economics and Statistics*, 76, 425–433.

Jayachandran, S. and I. Kuziemko (2011): "Why do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India," *The Quarterly Journal of Economics*, 126, 1485–1538.

Kelchtermans, S. and R. Veugelers (2013): "Top Research Productivity and its Persistence: Gender as a Double-Edged Sword," *The Review of Economics and Statistics*, 95, 273–285.

Kenney, G. M. and D. A. Wissoker (1994): "An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers," *The American Economic Review*, 84, 674–683.

Knepper, M. (2018): "When the Shadow Is the Substance: Judge Gender and the Outcomes of Workplace Sex Discrimination Cases," *Journal of Labor Economics*, 36, 623–664.

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

Kreisman, D. and M. A. Rangel (2015): "On the Blurring of the Color Line: Wages and Employment for Black Males of Different Skin Tones," *The Review of Economics and Statistics*, 97, 1–13.

Kuhn, P. and K. Shen (2013): "Gender Discrimination in Job Ads: Evidence from China," *The Quarterly Journal of Economics*, 128, 287–336.

Lang, K. and M. Manove (2011): "Education and Labor Market Discrimination," *The American Economic Review*, 101, 1467–1496.

Lange, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25, 1–35.

Leonard, J. S., D. I. Levine, and L. Giuliano (2010): "Customer Discrimination," *The Review of Economics and Statistics*, 92, 670–678.

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.

——— (2006): ""Friend or Foe?" A Natural Experiment of the Prisoner's Dilemma," *The Review of Economics and Statistics*, 88, 463–471.

Mechtenberg, L. (2009): "Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages," *The Review of Economic Studies*, 76, 1431–1459.

Miller, A. R. and C. Segal (2012): "Does Temporary Affirmative Action Produce Persistent Effects? A Study of Black and Female Employment in Law Enforcement," *The Review of Economics and Statistics*, 94, 1107–1125.

Mobius, M. M. and T. S. Rosenblat (2006): "Why Beauty Matters," *The American Economic Review*, 96, 222–235.

Nardinelli, C. and C. Simon (1990): "Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball," *The Quarterly Journal of Economics*, 105, 575–595.

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.

Neggers, Y. (2018): "Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India," *American Economic Review*, 108, 1288–1321.

Neumark, D., R. J. Bank, and K. D. V. Nort (1996): "Sex Discrimination in Restaurant Hiring: An Audit Study," *The Quarterly Journal of Economics*, 111, 915–941.

Neumark, D. and W. A. Stock (1999): "Age Discrimination Laws and Labor Market Efficiency," *Journal of Political Economy*, 107, 1081–1125.

Oettinger, G. S. (1996): "Statistical Discrimination and the Early Career Evolution of the Black- White Wage Gap," *Journal of Labor Economics*, 14, 52–78.

Ondrich, J., S. Ross, and J. Yinger (2003): "Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses from Black Customers?" *The Review of Economics and Statistics*, 85, 854–873.

Oreopoulos, P. (2011): "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes," *American Economic Journal: Economic Policy*, 3, 148–171.

Park, K. H. (2017): "Do Judges Have Tastes for Discrimination? Evidence from Criminal Courts," *The Review of Economics and Statistics*, 99, 810–823.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011): "Strike Three: Discrimination, Incentives, and Evaluation," *The American Economic Review*, 101, 1410–1435.

Persico, N., A. Postlewaite, and D. Silverman (2004): "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height," *Journal of Political Economy*, 112, 1019–1053.

Plug, E., D. Webbink, and N. Martin (2014): "Sexual Orientation, Prejudice, and Segregation," *Journal of Labor Economics*, 32, 123–159.

Price, J. and J. Wolfers (2010): "Racial Discrimination Among NBA Referees," *The Quarterly Journal of Economics*, 125, 1859–1887.

Rapaport, C. (1995): "Apparent Wage Discrimination when Wages are Determined by Nondiscriminatory Contracts," *The American Economic Review*, 85, 1263–1277.

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

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

Rubinstein, Y. and D. Brenner (2014): "Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labour Market Discrimination," *The Review of Economic Studies*, 81, 389–425.

Shayo, M. and A. Zussman (2011): "Judicial Ingroup Bias in the Shadow of Terrorism," *The Quarterly Journal of Economics*, 126, 1447–1484.

——— (2017): "Conflict and the Persistence of Ethnic Bias," *American Economic Journal: Applied Economics*, 9, 137–65.

Szymanski, S. (2000): "A Market Test for Discrimination in the English Professional Soccer Leagues," *Journal of Political Economy*, 108, 590–603.

Tootell, G. M. B. (1996): "Redlining in Boston: Do Mortgage Lenders Discriminate Against Neighborhoods?" *The Quarterly Journal of Economics*, 111, 1049–1079.

Weber, A. and C. Zulehner (2014): "Competition and Gender Prejudice: Are Discriminatory Employers Doomed to Fail?" *Journal of the European Economic Association*, 12, 492–521.

Wolfers, J. (2006): "Diagnosing Discrimination: Stock Returns and Ceo Gender," *Journal of the European Economic Association*, 4, 531–541.

Wozniak, A. (2015): "Discrimination and the Effects of Drug Testing on Black Employment," *The Review of Economics and Statistics*, 97, 548–566.

Inaccurate Statistical Discrimination: An Identification Problem

J. Aislinn Bohren, Kareem Haggag, Alex Imas, and Devin G. Pope.

December 16, 2022

Supplemental Material: Qualtrics Surveys

This file includes the Qualtrics survey used in the MTurk worker math trivia task, followed by the survey used in the MTurk employer hiring task. Survey block titles (not shown to participants) are in bold and underlined.

CHICAGO BOOTH

The University of Chicago Booth School of Business

Intro

Thank you for participating in this survey!

The survey has two parts. In the first part you will answer some very basic demographics questions. In the second part you will answer 50 multiple-choice math questions.

We are interested in determining how many of these math questions you can get right without any help. So please **do not** use a calculator or look up the answers online, but rather just do your best. The number of questions you answer correctly will not affect your payment in any way.

Demographics

Please answer the personal profile questions below:

What is your favorite color?

What is your favorite movie?

Do you prefer coffee or tea?

- ○ Tea

- ○ Coffee

What is your age?

What is your gender?

- ○ Female
- ○ Male

What is your favorite subject in high school?

What is your favorite sport?

Math

Q1. What is the square root of 289?

- ○ 17
- ○ 19
- ○ 15
- ○ 21

Q2. 4-8*9/2 = ?

- ○ -6
- ○ -32
- ○ -18
- ○ -12

Q3. 3^5 = ?

- ○ 243
- ○ 405

- ○ 729
- ○ 81

Q4. 5*6*7 = ?

- ○ 233
- ○ 210
- ○ 240
- ○ 180

Q5. What is the reduced form of the fraction 70/42?

- ○ 7/5
- ○ 14/10
- ○ 5/3
- ○ 10/6

Q6. What is the cubic root of 64?

- ○ 4
- ○ 6
- ○ 5
- ○ 3

Q7. (4+5)/5 = ?

- ○ 6.25
- ○ 1
- ○ 1.8
- ○ 5

Q8. x+2 < 18/3. Which of the following is necessarily **false**?

- ○ x > 4
- ○ x < 3
- ○ x > 3
- ○ x < 4

Q9. x^5 * x^8 = ?

- x^11
- x^14
- x^13
- x^12

Q10. Which of the following is approximately equal to 0.833?

- 5/6
- 4/5
- 6/7
- 3/4

Q11. x=5, y=6, z=7, then what is xy/(z-4)?

- 8
- 10
- 6
- 4

Q12. Which of the following is the closest integer to 45/7?

- 6
- 5
- 7
- 8

Q13. Which of the following is an integer multiple of 9?

- 3618
- 3619
- 3617
- 3620

Q14. 10/5+34-4 = ?

- 32
- 34
- 30
- 36

Q15. (x-1)*(x^2-4)=0, then which of the following **cannot** be x?

- 2
- -2
- -1
- 1

Q16. What is the square root of 196?

- 12
- 13
- 15
- 14

Q17. 5-6/(18/9) = ?

- 2
- -2
- -0.5
- 0.5

Q18. (y+9)*(y^2-121)=0, then which of the following **cannot** be y?

- 11
- 9
- -9
- -11

Q19. Which of the following is an integer multiple of 11?

- 133
- 130
- 132
- 131

Q20. 5+6+7+8+9+10 = ?

- 45
- 51
- 42
- 48

Q21. What is the binary form of 7?

- 101
- 100
- 111
- 110

Q22. 35/7+1 = ?

- 6
- 4
- 7
- 5

Q23. 24/4/3 = ?

- 4
- 3
- 1
- 2

Q24. Which of the following in an integer multiple of 4?

- 66
- 62

- 56
- 74

Q25. Which of the following is **not** a prime number?

- 4
- 2
- 3
- 5

Q26. 2*3*4*5 = ?

- 720
- 24
- 240
- 120

Q27. 6^3=?

- 216
- 432
- 36
- 128

Q28. (4*2+7*8)/4=?

- 20
- 24
- 16
- 12

Q29. Which of the following is a prime number?

- 23
- 27
- 21

- 25

Q30. 16 < x+8 < 26. Which of the following could x be?

- 23
- 18
- 13
- 8

Q31. 45+3-1 = ?

- 48
- 46
- 47
- 49

Q32. x^6 + x^6 = ?

- x^12
- x^36
- (2x)^6
- 2x^6

Q33. Which of the following fractions cannot be further reduced?

- 7/35
- 46/2
- 3/5
- 3/6

Q34. Which of the following numbers has an integer square root?

- 40
- 48
- 32
- 36

Q35. 5*(7+3)+5-4 = ?

- 51
- 55
- 39
- 32

Q36. Which of the following is **not** a factor of 30?

- 3
- 5
- 2
- 4

Q37. x^6 / x^4 = ?

- x^24
- x^10
- x^2
- x^(2/3)

Q38. 56/8 = ?

- 6
- 5
- 7
- 8

Q39. 2^4 - 3^3 = ?

- 11
- 9
- -11
- -9

Q40. (18+19+20)/3 = ?

- 20
- 21
- 19
- 18

Q41. Twenty **cannot** be divided by which of the following?

- 5
- 3
- 2
- 4

Q42. 4+8+12+16 = ?

- 40
- 20
- 25
- 45

Q43. (x^5)^3 = ?

- 5x^3
- 3x^5
- x^15
- x^8

Q44. Which of the following is the correct factorization of 36?

- 4 * 9
- 2^2 * 3^2
- 4 * 3^2
- 2^2 * 9

Q45. 3^2*2 = ?

- 18
- 42

- 81
- 24

Q46. -2*(-3-8) = ?

- -14
- 14
- 22
- -22

Q47. Which of the following is an integer multiple of 5?

- 44
- 46
- 43
- 45

Q48. x^4 = 81. What is x?

- 9
- 20.5
- 3
- 6

Q49. 76/4 = ?

- 18
- 19
- 17
- 20

Q50. Which of the following is negative?

- 2^2
- (-2)^2
- (-2)^3

○ 2^3

Final

Thank you for your participation. In addition to your base payment, we may put a small bonus into your account sometime in the next few weeks. Who receives the bonus payment is determined by a different experiment that we are doing and is unrelated to how well you did in the task. Please just think of it as an additional appreciation for your efforts.

CHICAGO BOOTH

The University of Chicago Booth School of Business

Introduction

Thank you for participating in this survey.

The survey has four parts. You will first answer some simple demographic questions. Then you will answer three sets of questions related to people's performance in math questions.

The survey will take approximately 20 minutes.

Please enter your M-Turk ID:

What is your gender?

- Male
- Female

What is your age?

Please indicate the highest level of education you have completed.

- Less than High School
- High School or equivalent
- Vocational/Technical School (2 year)

- ○ Some College
- ○ College Graduate (4 year)
- ○ Master's Degree (MS)
- ○ Doctoral Degree (PhD)
- ○ Professional Degree (MD, JD, etc.)
- ○ Other

Setup

We recently paid many people to answer 50 math questions each. Here are some examples of the types of math questions we asked:

Question 1: What is the square root of 289?
Choices: 15, 17, 19, 21

Question 2: 4-8*9/2 = ?
Choices: -6, -12, -18, -32

Question 3: What is the reduced form of the fraction 70/42?
Choices: 5/3, 10/6, 7/5, 14/10

Question 4: x^5 * x^8 = ?
Choices: x^11, x^12, x^13, x^14

Question 5: What is the binary form of 7?
Choices: 100, 101, 110, 111

On average, participants answered 36.95 out of 50 questions correctly.

Today, you are going to be an employer. You will hire one of the people who answered our math questions. The person you hire will be given a bonus (the wage that you choose to pay them) and in return you will receive money based on how many of the math questions they answered correctly.

Specifically, we are going to provide you with the profiles of 20 people (potential

Specifically, we are going to provide you with the profiles of 20 people (potential employees) who answered our math questions. For each of the 20 people that we present, you will indicate what is the highest wage (between 0 and 50 cents) you would be willing to pay that person. In return, you will be paid 1 cent for every question that the person you end up hiring answered correctly.

After you indicate the highest wage you would be willing to give to each employee, we will randomly draw a number between 0 and 50. If the wage you chose for the employee is equal to or higher than the randoly-drawn number, then that employee will receive the random number as a bonus, and you will receive a profit equal to the number of correct answers given by the individual minus the random number that was drawn. If the highest wage you were willing to pay the individual is lower than the random number, you will not hire the employee and neither you nor the employee will receive a bonus.

Let's walk through an example of how this works. Below is an example of a potential employee profile that you might see:

Country:	United States
Gender:	Female
Age:	63
Favorite High School Subject:	English
Favorite Sport:	Gymnastics
Favorite Color:	Sea Green
Favorite Movie:	Overboard
Prefers Coffee/Tea:	Tea

We will ask you the highest amount you would be willing to pay this employee. Let's imagine that you say you would be willing to pay this employee 40 cents.

We will then select a random number between 0 and 50. Let's say the randomly-selected number is 20. Because the highest wage you are willing to pay that person is more than 20, you will "hire" this person and they will receive 20 cents. You will then be paid based on the number of correct answers this person gave. If the person answered 30 questions correctly, you will be paid 10 cents (30-20). If the person answered 10 questions correctly, you will be paid -10 cents (10-20).

Imagine instead that the randomly-drawn number is 45. Then you will not "hire" the person and neither you nor the person will receive a bonus.

In today's task, you will actually only hire 1 person. After you decide the most you would be willing to pay to each of the 20 people we present, we will randomly select one profile to use as the actual hiring decision. We will then draw the random number between 0 and 50 and pay you the profit you've earned for that profile and pay the wage to the person whose profile you pick. We are going to automatically give you a $0.50 bonus in addition to what money you make with your hiring decision (so that there is no way you end up owing us any money after doing this task).

Just to make sure you understand, imagine you saw a profile and entered **43** as the highest amount you would be willing to pay. Now imagine the random number generated was **18** and the individual answered **10** questions correctly.

How many cents would you have to pay the individual?

How many cents would you be paid based on the individual's performance (before subtracting the wage you have to pay the individual)?

Suppose instead that you had reported **15** as the highest wage you would pay, and everything else stayed the same:

How many cents would you have to pay the individual?

How many cents would you be paid based on the individual's performance (before subtracting the wage you have to pay the individual)?

Hidden Generator

Required

You have completed ${lm://Field/1} of 20 required profiles.

Please indicate the **highest wage** you would be willing to pay this employee in the text box below.

Enter the highest wage you would be willing to pay this individual (between 0 and 50 cents):

Prediction

Thank you for completing part 2 of 4 of this survey. As promised, we will randomly select one profile and pay you your $0.50 bonus plus whatever money you make

based on the hiring of the randomly-selected profile.

For the third part of this survey, please answer the six questions below. Please remember that people answered **36.95** questions correctly on average.

On average, how many math questions out of 50 do you think **women** answered correctly?

On average, how many math questions out of 50 do you think **men** answered correctly?

On average, how many math questions out of 50 do you think **people from the United States** answered correctly?

On average, how many math questions out of 50 do you think **people from India** answered correctly?

On average, how many math questions out of 50 do you think **people below or at the age of 33** answered correctly?

On average, how many math questions out of 50 do you think **people above the age of 33** answered correctly?

Thank you for completing part 2 of 4 of this survey. As promised, we will randomly select one profile and pay you your $0.50 bonus plus whatever money you make based on the hiring of the randomly-selected profile.

For the third part of this survey, please answer the six questions below. Please remember that people answered **36.95** questions correctly on average.

You have the chance to earn a significant bonus if you answer these questions correctly. We will randomly pick one question and pay you $5 minus your deviation from the correct answer. For example, if your answer for the randomly picked question is 40 and the truth is 37, then you will get a $2 bonus. You cannot receive a negative bonus. So, please answer the questions as carefully as possible so that you can potentially win a large bonus.

On average, how many math questions out of 50

do you think **women** answered correctly?	
On average, how many math questions out of 50 do you think **men** answered correctly?	
On average, how many math questions out of 50 do you think **people from the United States** answered correctly?	
On average, how many math questions out of 50 do you think **people from India** answered correctly?	
On average, how many math questions out of 50 do you think **people below or at the age of 33** answered correctly?	
On average, how many math questions out of 50 do you think **people above the age of 33** answered correctly?	

Truth

Here are the correct answers for the 6 questions you have answered above. On average:

- Women got **35.28** questions right.
- Men got **38.32** questions right.
- People from the U.S. got **37.14** questions right.
- People from India got **36.58** questions right.
- People below or at the age of 33 got **37.10** questions right.
- People above the age of 33 got **36.79** questions right.

Now that you have learned those facts, we would like you to work on 10 more profiles.

As before, after you finish working on those 10 additional profiles, we will randomly select one profile and randomly select a number between 0 and 50. If your highest wage is more than the randomly-selected number, we will pay you the profit you've earned for that profile as a bonus and pay the wage to the person who answered the math questions.

Extra

You have completed ${lm://Field/1} of 10 additional profiles.

Please indicate the **highest wage** you would be willing to pay this employee in the text box below.

Enter the highest wage you would be willing to pay this individual (between 0 and 50 cents):

Final

Thank you for your participation. We will calculate your bonus based on the rules specified in each part above, and pay the bonus to your account within a week.

If you have any additional comments about this survey, please provide them below. (Optional)