Inaccurate Statistical Discrimination: An Identification Problem

Online Appendix

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

August 3, 2023

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). $\square$

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. $\square$

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 1 (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:*

$$u_g = \left(\frac{\eta_g}{\tau_g + \eta_g}\right) s_g + \left(\frac{\tau_g}{\tau_g + \eta_g}\right) \mu_g. \quad (1)$$

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

$$u'_g = \left(\frac{\eta_g}{\tau_g + \eta_g}\right) s_g + \left(\frac{\tau_g}{\tau_g + \eta_g}\right) \mu_g. \tag{2}$$

In contrast, the the true preference parameter satisfies

$$u_g = \left(\frac{\hat{\eta}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) s_g + \left(\frac{\hat{\tau}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) \hat{\mu}_g. \tag{3}$$

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

$$\mu_g = \left(\frac{\tau_g + \eta_g}{\tau_g}\right) \left[\left(\frac{\hat{\eta}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) s_g - \left(\frac{\eta_g}{\tau_g + \eta_g}\right) s_g + \left(\frac{\hat{\tau}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) \hat{\mu}_g\right]. \tag{4}$$

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

$$u'_F = \left(\frac{\eta_F}{\tau_F + \eta_F}\right) s_F + \left(\frac{\tau_F}{\tau_F + \eta_F}\right) \mu_F \geq u_F. \tag{5}$$

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

$$u_g = \left(\frac{\hat{\eta}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) s_g + \left(\frac{\hat{\tau}_g}{\hat{\tau}_g + \hat{\eta}_g}\right) \hat{\mu}_g. \tag{6}$$

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

$$\frac{\hat{\tau}_g + x_i \hat{\eta}_g}{x_i \hat{\eta}_g} u_g - \frac{\hat{\tau}_g}{x_i \hat{\eta}_g} \hat{\mu}_g = s_{g,i}. \tag{7}$$

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

$$\left(\frac{\hat{\tau}_g}{\hat{\eta}_g} + x_i\right) u_g = x_i s_{g,i} + \frac{\hat{\tau}_g}{\hat{\eta}_g} \hat{\mu}_g. \tag{8}$$

Subtracting Eq. (8) evaluated at x_2 from Eq. (8) evaluated at x_1 and solving for u_g identifies the evaluator's preferences as

$$u_g = \frac{x_1 s_{g,1} - x_2 s_{g,2}}{x_1 - x_2}. \tag{9}$$

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. (9). 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

$$\frac{\hat{\tau}_g + x\hat{\eta}_g}{x\hat{\eta}_g} u_g - \frac{\hat{\tau}_g}{x\hat{\eta}_g} \hat{\mu}_g = \frac{\hat{\tau}'_g + x\hat{\eta}'_g}{x\hat{\eta}'_g} u_g - \frac{\hat{\tau}'_g}{x\hat{\eta}'_g} \hat{\mu}'_g \tag{10}$$

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

$$\frac{\hat{\tau}_g(u_g - \hat{\mu}_g)}{\hat{\eta}_g} = \frac{\hat{\tau}'_g(u_g - \hat{\mu}'_g)}{\hat{\eta}'_g} \tag{11}$$

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. (11) is independent of x, any $\theta' = (u_g, \hat{\mu}'_g, \hat{\tau}'_g, \hat{\eta}'_g)_{g\in\{M,F\}}$ that satisfies Eq. (10) 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

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

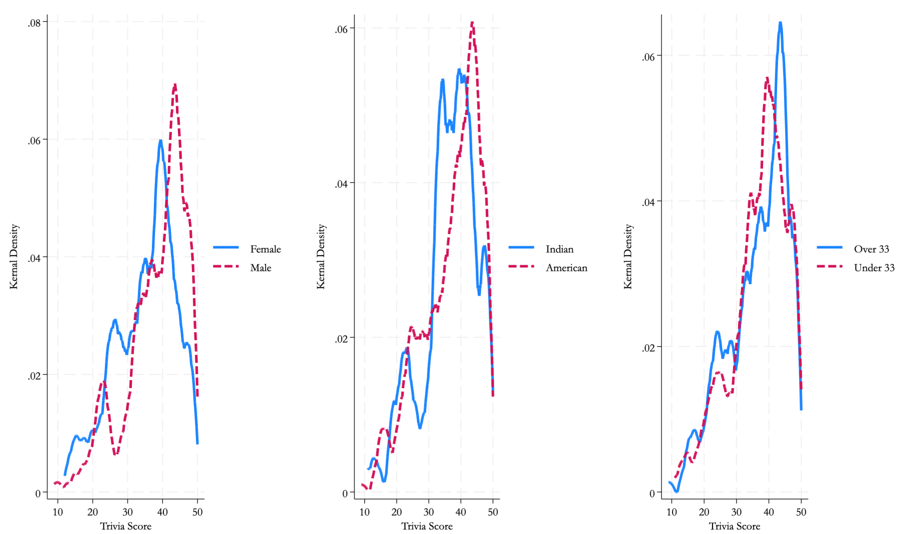

Figure B2. Kernel Densities of Beliefs about Differences by Group (Within-Employer)

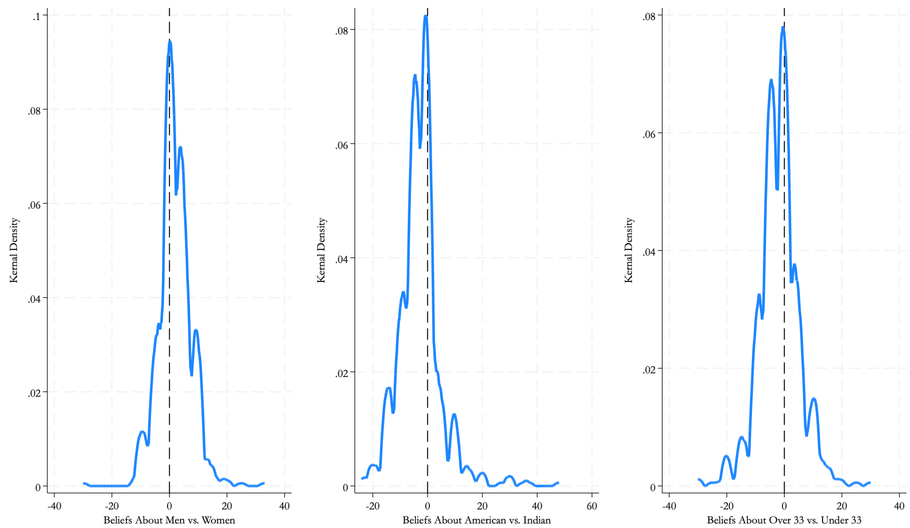

Table B1. Summary Statistics

	Total (1)	Male (2)	Female (3)	US (4)	India (5)	Under 33 (6)	Over 33 (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).

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 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 1.31 (p=0.260).

Table B6. Beliefs about Productivity by Employee Characteristics, Trimmed

	Group 1	**Group 2**	**Diff.**	**p-val**
	(1)	(2)	(3)	(4)
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).

Appendix C. Literature Survey Details

In this section, we further discusses the seven papers in Table 1 that tested whether inaccurate beliefs could be driving discrimination, then we present our methodology for the literature survey in Section 2.

C.1 Discussion of Inaccurate Belief Literature

Of the seven papers that consider inaccurate beliefs in identification, List (2004); Hedegaard and Tyran (2018); Mobius and Rosenblat (2006); Beaman, Chattopadhyay, Duflo, Pande, and Topalova (2009) 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. Finally, Beaman et al. (2009) study the effects of gender quotas on perceptions of leader effectiveness. They find that male villagers in control villages evaluate hypothetical male (vs. female) leaders as significantly more effective, but this evaluation gap disappears in villages with quotas for female leaders. Moreover, they present evidence that male leaders do not outperform female leaders, suggesting this evaluation gap may be driven by inaccurate beliefs.

Agan and Starr (2017); Arnold, Dobbie, and Yang (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, Coffman, Gennaioli, and Shleifer (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.[1]

C.2 Methodology

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 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

[1] 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**

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.[2]

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

[2]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.

identity. 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.

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%

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.

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

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.

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

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.

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

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

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

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

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

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