

# Inaccurate Statistical Discrimination

Bohren, Haggag, Imas, Pope (2019)

August 2019

- Discrimination has been widely studied in economics and other disciplines. In addition to identifying evidence of discrimination, economists often categorize the source of discrimination as either taste-based or statistical.
- A further categorization is important and needed. Specifically, in many situations economic agents may have inaccurate beliefs about the expected productivity or performance of a social group.
- Accurate (based on correct beliefs) and inaccurate (based on incorrect beliefs) statistical discrimination

- Taste-based discrimination (Becker 1957): an individual or firm has animus towards members of a particular group, choose to discriminate against them because he receives disutility from providing services to or interacting with members of the group.
- Statistical discrimination: discrimination may occur against members of a particular group because productivity is unobserved and the group is perceived to have a lower average productivity (Phelps 1972; Arrow 1973) or a different variance.
- Designing an effective policy intervention to reduce discrimination crucially depends on the source of discrimination.
- Welfare/efficiency analyses differ for each type of discrimination.

# Inaccurate Statistical Discrimination

- In many situations, an individual's beliefs about the productivity of different social groups may be inaccurate.
- We refer to discrimination that stems from inaccurate beliefs as inaccurate statistical discrimination.
- Policy design and welfare analysis: if discrimination stems from inaccurate beliefs, an effective policy response could be to provide individuals with information about the correct distributions

- By not allowing for the possibility of inaccurate beliefs, we demonstrate that the discrimination literature may mistakenly interpret evidence of inaccurate statistical discrimination as taste-based and/or accurate statistical discrimination.
- We conducted a systematic survey of the economics literature on discrimination in order to determine:
  - How often papers seek to distinguish between taste-based and belief-based (statistical) sources of discrimination
  - How often papers seek to distinguish between accurate and inaccurate beliefs for belief-based sources of discrimination.

Table 1: Summary of Literature Survey on Discrimination

	All: 1990 - 2018		Recent: 2014 - 2018	
	# Papers	Percent	# Papers	Percent
Total Papers	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	5	4.8%	2	6.5%
Measure beliefs	6	5.7%	3	9.7%

# Hiring Experiment

- Stylized hiring experiment that allows us to perform an accounting exercise typical of the outcomes-based tests common in the literature and to elicit relevant beliefs.
- We show that actual beliefs are inaccurate, violating the rational expectations assumption typically made in the statistical discrimination literature.
- We also test a simple intervention to correct inaccurate beliefs: we provide individuals with information on the distributions of productivities by groups and conduct an additional hiring task. This demonstrates the effectiveness of this type of informational intervention.

- Two separate, pre-registered surveys:
  - A work task (math quiz) performed by 589 Amazon Mechanical Turk subjects (MTurkers), who comprise the prospective workers for the second survey
  - A hiring task in which each of 577 different MTurkers, who comprise the employers, stated a wage (willingness to pay) for 20 prospective worker profiles.
  - The second survey also contains a belief elicitation and an information intervention followed by a second hiring task.



# Survey 1 (Work Task)

- The purpose of the first survey was to create a bank of "workers" who could be hired by the "employers" in the second survey.
- 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 percent or higher approval rate for these tasks.
- Math Questions and Demographics
- 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.

# Survey 1 (Work Task)

- In contrast to correspondence studies, we did not employ deception at any point. 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.
  - Demographic questions: relevant for animus and/or beliefs about productivity (e.g. age, gender, and nationality), as well as other irrelevant information
  - The irrelevant information serves as a placebo test.
- 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.

# Survey 1 (Work Task)

- The purpose of the first survey was to create a bank of "workers" who could be hired by the "employers" in the second survey.
- 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 percent or higher approval rate for these tasks.
- Math Questions and Demographics
- 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.

# Survey 2

- We recruited 577 different MTurk subjects on February 26, 2018. We used the same hiring criteria as the first survey.
- 20-Minute Survey about Decision-Making”
- We paid 2 USD.
- Subjects were first asked to report their gender, age, and education level.

# First Hiring Task

- We informed subjects that we had previously paid other subjects ("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. 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.
- In return, they would receive a payment based on how many questions their hired worker answered correctly.
- the mechanism (Becker-DeGroot-Marschak) used to assign payment.

# 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

- Subjects were randomly assigned to one of two different conditions: an incentivized or un-incentivized belief elicitation.
- 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."
- In the incentivized condition, prior to the six questions, subjects were told that they could earn a significant bonus for an accurate prediction.

# Information Intervention and 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).
- 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.



- There is variation in this timing across groups.
- Another large difference between the U.S. and India samples is the average age of participants.
- The Indian sample also skews more male than the U.S. sample and is more likely to have a college education or above
- While we primarily focus on simple comparisons between each demographic group, these observed differences motivate our use of multivariate regressions as well.

# Experimental Results

- We did find evidence of discrimination with respect to two out of three group identities: gender and nationality, but not age.

	(1)	(2)	(3)
	Group	Group	Diff.
	1	2	[(1)-(2)]
<b>Panel A: Employers' Wage WTP, by Employee Characteristics</b>			
Gender (1 = Male , 2 = Female)	31.90 (12.07)	30.85 (12.23)	1.05
Country (1 = US , 2 = India)	30.71 (12.20)	32.85 (11.95)	-2.14
Age (1 = Under 33 , 2 = Over 33)	31.67 (12.00)	31.14 (12.33)	0.54
Placebo (1 = Prefer Coffee , 2 = Prefer Tea)	31.22 (12.32)	31.74 (11.89)	-0.52
<b>Panel B: Employee Productivity, by Employee Characteristics</b>			
Gender (1 = Male , 2 = Female)	38.30 (8.55)	34.98 (8.73)	3.32
Country (1 = US , 2 = India)	37.01 (8.93)	36.36 (8.49)	0.65
Age (1 = Under 33 , 2 = Over 33)	36.96 (8.62)	36.60 (8.98)	0.37
Placebo (1 = Prefer Coffee , 2 = Prefer Tea)	36.64 (8.77)	37.03 (8.82)	-0.40

# Experimental Results

- regression results that simultaneously control for all demographic variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Female	-1.05*** (0.25)				-0.67*** (0.24)	-0.80*** (0.19)
Indian		2.14*** (0.29)			1.98*** (0.29)	2.00*** (0.25)
Over 33			-0.54** (0.26)		0.07 (0.26)	0.32 (0.22)
Placebo: Prefers Tea				0.52** (0.24)	0.39* (0.24)	0.37** (0.18)
N	11,540	11,540	11,540	11,540	11,540	11,540
$R^2$	0.00	0.01	0.00	0.00	0.01	0.49
DepVarMean	31.90	30.71	31.67	31.22	30.18	30.18
Employer FE?	No	No	No	No	No	Yes

# Experimental Results

- The gap in average wages for men and women was lower than the gap in average performance (1.05 points versus 3.32 points). Therefore, if we used this standard approach 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. Further, we would conclude that the remaining 2.17 point difference in performance suggests 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 both the -2.14 point disparity in wages and the 0.65 point difference in performance suggest taste-based discrimination against Americans.

- this standard approach naively assumes that employers have accurate beliefs about group-level performance differences. To investigate this assumption, we elicited beliefs about performance.

# Experimental Results

- Employer beliefs about the group-level average performance (math quiz scores) that can be compared to the actual performance reported before.

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

# Experimental Results - Inaccurate Statistical Discrimination

- Nationality: the wage gap is -2.14 points and the performance gap is 0.65 points, whereas the belief gap is -2.72 points. Thus, the entire wage gap is explained by inaccurate beliefs. The remaining 0.58 point (2.72 minus 2.14) difference between the belief and wage gaps suggests taste-based discrimination against Indians.
- Gender: The wage gap is 1.05 points, the performance gap is 3.32 points, and the belief gap is 1.89 points. The wage gap is still fully accounted for by beliefs, but the gap between performance and wages is partly explained by inaccurate beliefs; the difference explained by a taste-based source shrinks from 2.17 to 0.84 points.
- Age: despite the minimal gap in wages and performance for age, employers believed that young workers outperform older ones. This suggests taste-based discrimination against younger workers.

- The role of inaccurate beliefs in explaining part of the wage gap suggests a policy intervention: inform employers of the correct distributions and provide them with the opportunity to set wages for a different set of potential workers to see if this induces a behavioral change.



# The effects of informational intervention

	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Post-Info	1.53*** (0.29)	1.60*** (0.27)	1.06*** (0.31)	1.28*** (0.29)	1.94*** (0.44)	2.39*** (0.36)
Female	-1.05*** (0.25)				-0.67*** (0.24)	-0.81*** (0.19)
Female X Post-Info	-0.64* (0.37)				-0.90** (0.37)	-1.00*** (0.29)
Indian		2.14*** (0.29)			1.98*** (0.29)	1.99*** (0.25)
Indian X Post-Info		-1.07*** (0.40)			-1.20*** (0.42)	-1.63*** (0.34)
Over 33			-0.54** (0.26)		0.07 (0.26)	0.30 (0.22)
Over 33 X Post-Info			0.41 (0.40)		0.14 (0.42)	-0.21 (0.30)
Prefers Tea				0.52** (0.24)	0.39* (0.24)	0.35** (0.18)
Prefers Tea X Post-Info				-0.08 (0.38)	0.06 (0.38)	-0.18 (0.27)
N	17,310	17,310	17,310	17,310	17,310	17,310
R <sup>2</sup>	0.01	0.01	0.00	0.00	0.01	0.48
DepVarMean	31.90	30.71	31.67	31.22	30.18	30.18
Employer FE?	No	No	No	No	No	Yes

- We compare the differences between the two hiring rounds (Post-Info), the differences between wages assigned to profiles of each worker demographic group (e.g. "Female"), and the resulting difference-in- differences (e.g. "Female X Post-Info").

# The effects of informational intervention

	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Post-Info	1.53*** (0.29)	1.60*** (0.27)	1.06*** (0.31)	1.28*** (0.29)	1.94*** (0.44)	2.39*** (0.36)
Female	-1.05*** (0.25)				-0.67*** (0.24)	-0.81*** (0.19)
Female X Post-Info	-0.64* (0.37)				-0.90** (0.37)	-1.00*** (0.29)
Indian		2.14*** (0.29)			1.98*** (0.29)	1.99*** (0.25)
Indian X Post-Info		-1.07*** (0.40)			-1.20*** (0.42)	-1.63*** (0.34)
Over 33			-0.54** (0.26)		0.07 (0.26)	0.30 (0.22)
Over 33 X Post-Info			0.41 (0.40)		0.14 (0.42)	-0.21 (0.30)
Prefers Tea				0.52** (0.24)	0.39* (0.24)	0.35** (0.18)
Prefers Tea X Post-Info				-0.08 (0.38)	0.06 (0.38)	-0.18 (0.27)
N	17,310	17,310	17,310	17,310	17,310	17,310
R <sup>2</sup>	0.01	0.01	0.00	0.00	0.01	0.48
DepVarMean	31.90	30.71	31.67	31.22	30.18	30.18
Employer FE?	No	No	No	No	No	Yes

- The results of the information intervention suggest that identifying inaccurate beliefs may have immediate policy implications for reducing discrimination.
- We argue that researchers studying motives of discrimination should in all situations consider and discuss the possibility of inaccurate beliefs and elicit beliefs when possible. By doing so, researchers are less likely to mistakenly assign motives of animus or accurate statistical discrimination when those may actually not be the actual intentions.