Self Justification and Belief Distortion

Christina Sun
University of California, Davis

ESA 2024

Overview

This project experimentally studies whether taste-based discrimination against a group can lead to biased beliefs about members from the group

Motivation

- Traditionally, discrimination in economics literature is classified into two mutually exclusive types: taste-based discrimination (Becker, 1957) and statistical discrimination (Phelps, 1972; Arrow, 1973)
- Recently, a new literature studies statistical discrimination stemming from inaccurate beliefs (e.g. Bohren et al., 2019; Bohren et al., 2023; Eyting, 2024; Mengel and Campo-Mercade, 2024)
- Little evidence exists on how inaccurate beliefs form in discrimination contexts
  - Eyting (2024) shows motivated reasoning can potentially lead to biased beliefs through biased information acquisition
  - I show self justification of discrimination and biased information processing/recall can lead to biased beliefs

Questions

- Does taste-based discrimination lead to biased beliefs about workers?
- Is this belief distortion a result of attempted justification/rationalization?
- What other mechanisms drive this belief distortion?
  - Biased Recall
  - Misuse of Statistics

Sources of Discrimination

The 2 Models of discrimination:

- They are thought of as distinct and mutually exclusive
- Taste-Based Discrimination | Statistical Discrimination

Sources of Discrimination

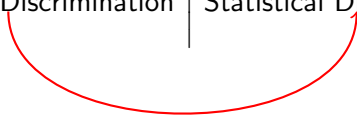
The 2 Models of discrimination:

- They are thought of as distinct and mutually exclusive
- Taste-Based Discrimination | Statistical Discrimination
- Or are they?

Sources of Discrimination

The 2 Models of discrimination:

- They are thought of as distinct and mutually exclusive
- Taste-Based Discrimination | Statistical Discrimination
- Or are they?



Experiment Overview

- A hiring game in which an employer makes binary hiring choice between pairs of workers and reports beliefs
- I shut down the possibility of statistical discrimination ex ante
- Neutral Frame vs. Race Frame
- Hiring + Belief vs. Belief Only

Design: Worker Survey

- Workers complete a timed math test and a real effort task
- They earn a total score between 0 to 30
- Worker groups are constructed so that the score distribution is the same across groups

Design: Hiring Experiment

1. Employers are informed about the workers
   - Neutral Treatment: Workers are labeled neutral terms (Purple, Orange)
   - Race Treatment: Workers are labeled with race (Black, Asian)
2. They report prior beliefs about the score distribution in the worker groups
3. Employers learn the true distribution
4. Employers report updated posteriors about group distributions
5. Employers make a series of binary hiring choices between randomly drawn worker pairs (not present in Belief Only)
6. Employers report beliefs about individual worker scores

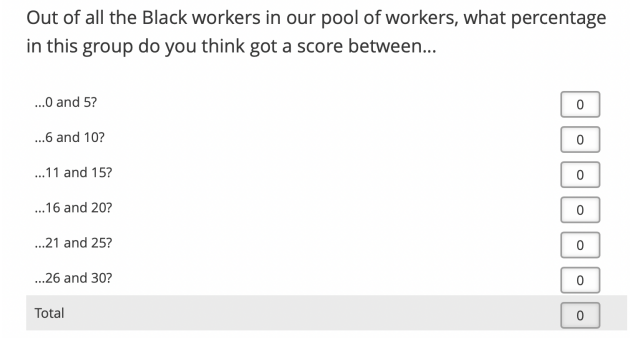
Incentives

Incentives:

- One decision is randomly selected to be paid
  - If a hiring decision is selected, the subject earns a bonus if they hired the worker with the higher score
  - If a belief is selected, the subject earns a bonus if they reported within a +- 2 window of the true score

Hiring Experiment Interface

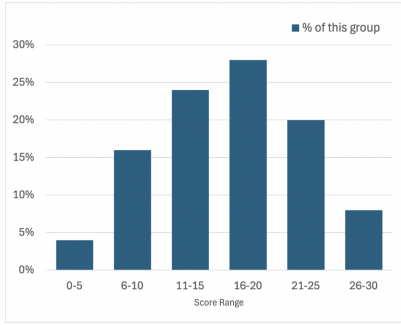
Eliciting prior beliefs for score distribution:



Hiring Experiment Interface

Information treatment by providing true distribution:

This graph shows the distribution of scores of all Black workers in our sample.

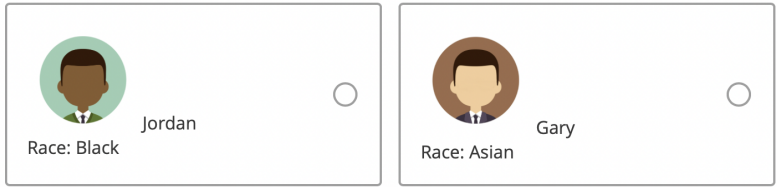


Out of all the Black workers in our pool of workers, what percentage of the group got a score between...

Hiring Experiment Interface

Hiring Decision in Race Frame:

Which of these two workers do you hire?



Hiring Experiment Interface

Belief elicitation in Race Frame:

What do you think is each worker's score? (out of 30)

Jordan

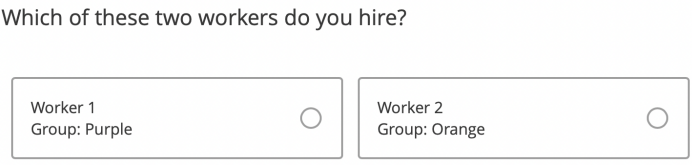
Race: Black

Gary

Race: Asian

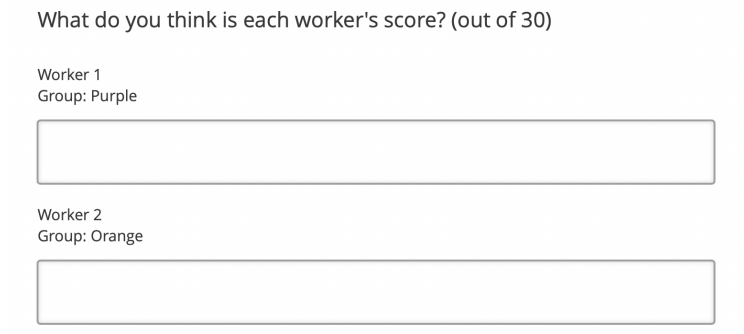
Hiring Experiment Interface

Hiring Decision in Neutral Frame:



Hiring Experiment Interface

Belief elicitation in Neutral Frame:



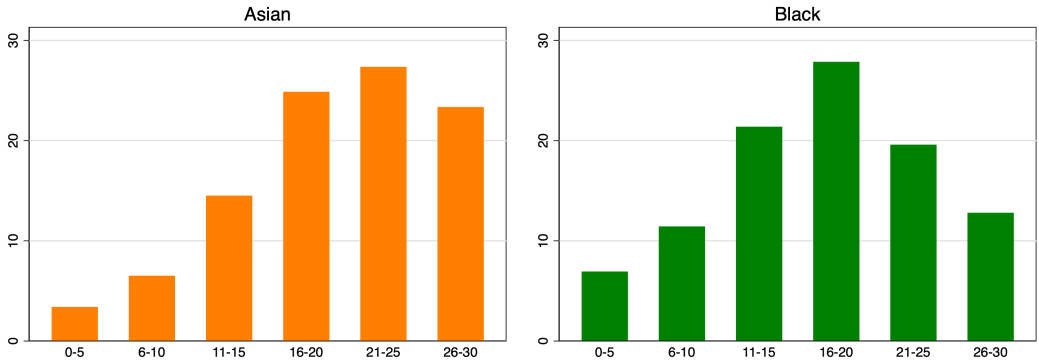
Implementation

- Experiment conducted on Prolific
- 392 subjects across 4 treatments

	Race	Neutral
Hiring	112	112
Belief Only	109	59

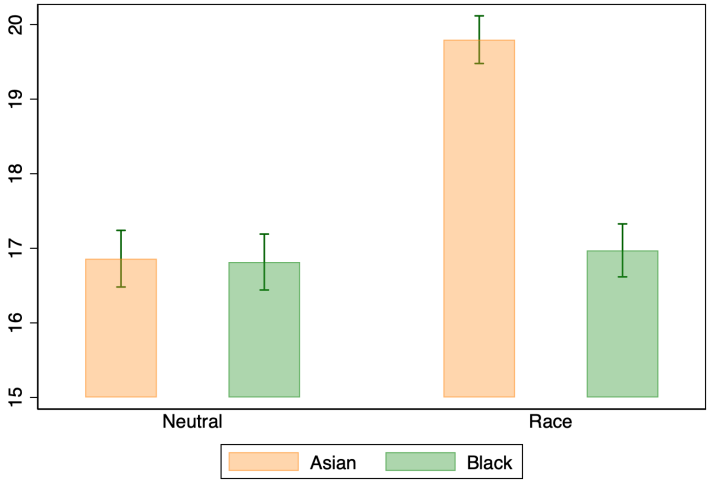
Results: Prior Beliefs for Group Distributions

Result 1(a): Employers hold different prior belief distributions for Asian and Black workers in the Race Frame, but not in Neutral Frame..



Results: Prior Means

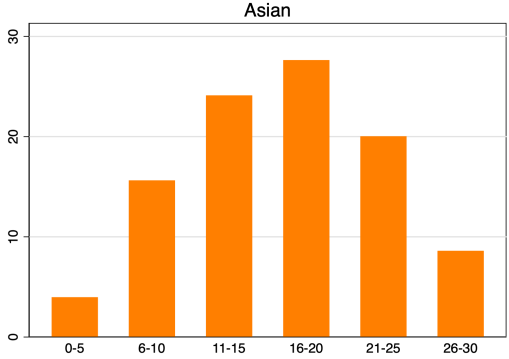
Prior Means in Neutral and Race Frames:



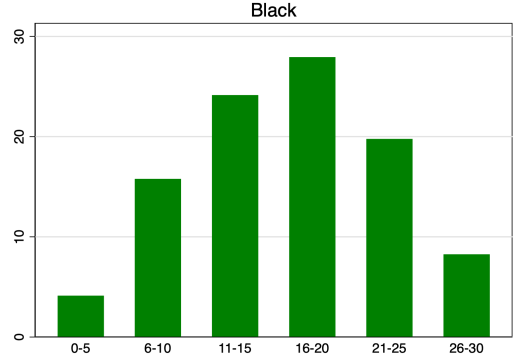
Results: Effect of Information Treatment

Result 1(b): Employers' beliefs responds to the true distribution. Information treatment is effective at equalizing posterior beliefs.

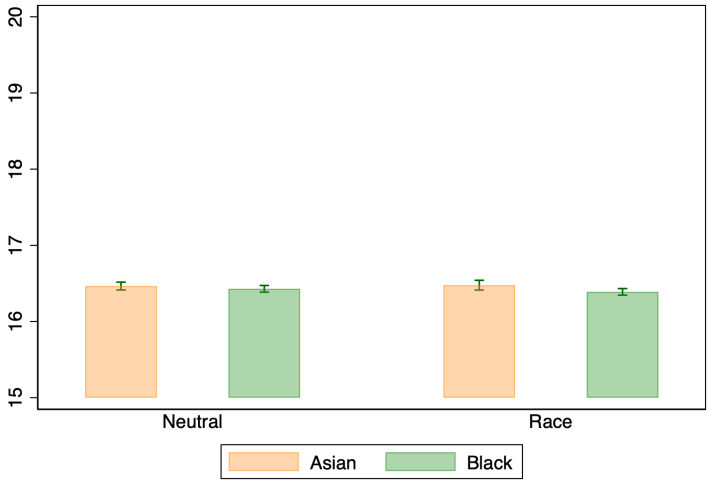
Asian



Black

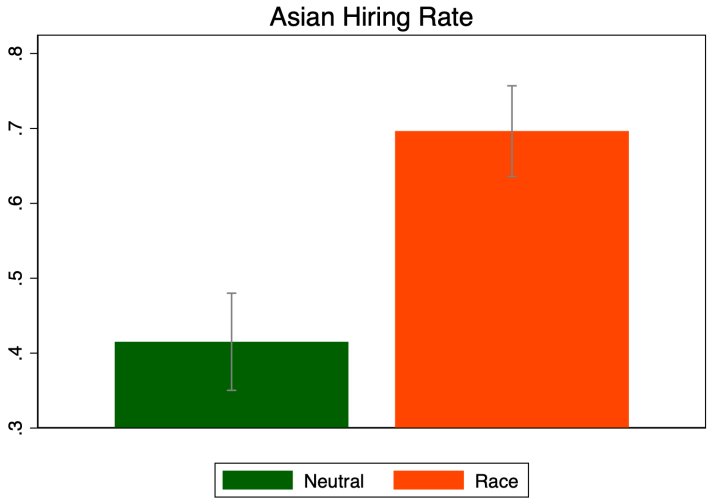


Results: Posterior Means



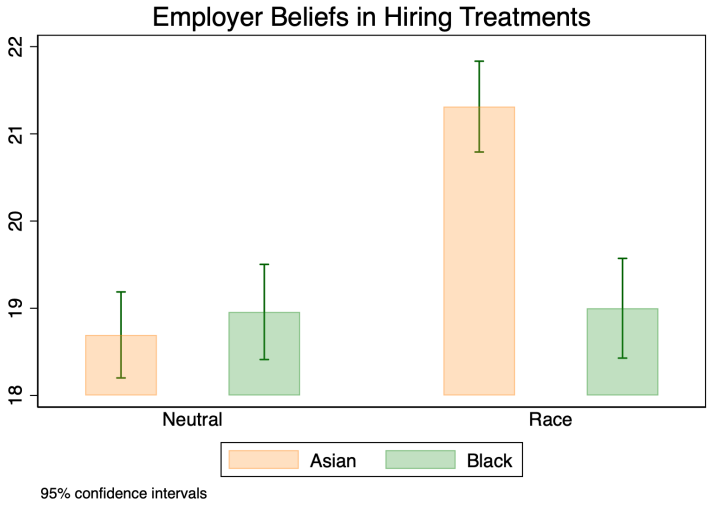
Results: Hiring Decision

Result 2: There exists significant discrimination against Black workers: significantly more employers hire Asian workers in Asian-Black pairs in the Race Frame.



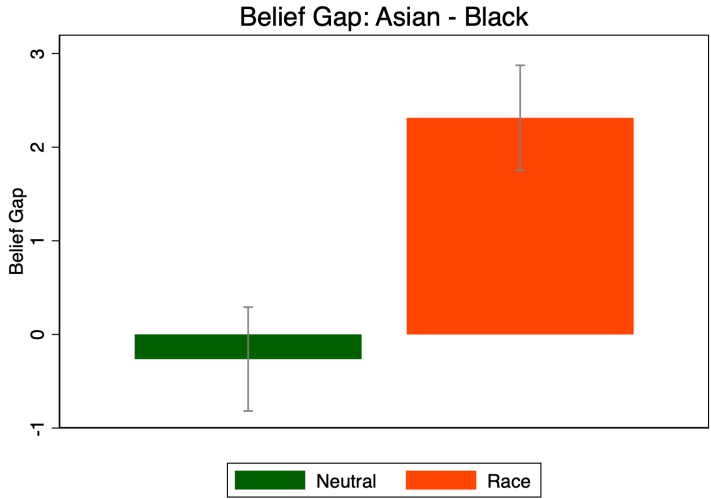
Results: Beliefs for Workers

Point beliefs for Asian and Black workers



Results: Beliefs for Workers

Result 3: There exists a substantial gap between the employers' beliefs for Asian and Black workers in the Race Frame.



Why is there a belief gap?

- The employer holds the same prior belief distribution for both Asians and Black groups
- Thus a rational employer should hold equal beliefs for randomly drawn Asian and Black workers
- Race Frame turns on race labels, which can work through
  1. Taste
  2. ~~Statistical~~
- Consistent with the interpretation that the presence of taste in Race Frame leads to belief distortions

Mechanisms

- Self justification/rationalization of discrimination
  - "I didn't hire Black not because I am racist, but because they have lower score"
- Motivated Memory
  - Bias against minority group can drive biased recall of score distributions
- Distorted statistical thinking
  - Employers could be thinking about drawing Asian workers from the right tail and Black workers from the left tail

Justification: Belief Only Treatment

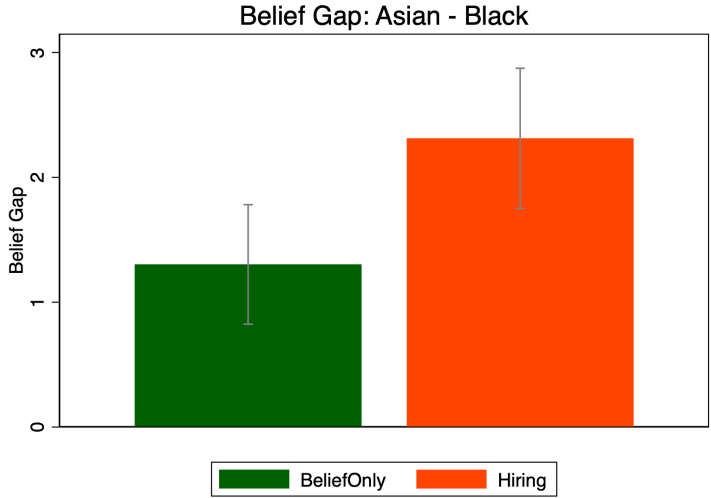
- Employers do not make hiring choices, only report beliefs
- Same as Hiring treatment otherwise
- Intuition: without hiring, less need to justify their choice in this treatment

Recall

- I ask employers to recall the score distributions for each worker group at the end of the experiment
- I compare the recall between Race and Neutral Frame in the Belief Only treatments
- This removes possible effects of self justification for hiring decisions leading to more biased recall

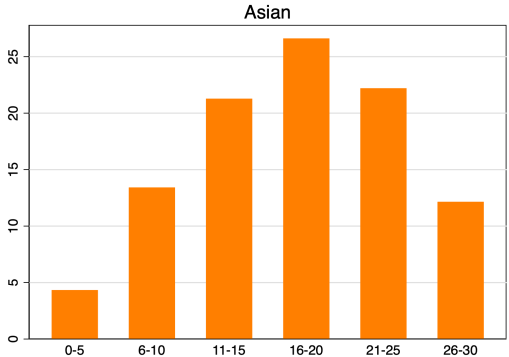
Results: Justification

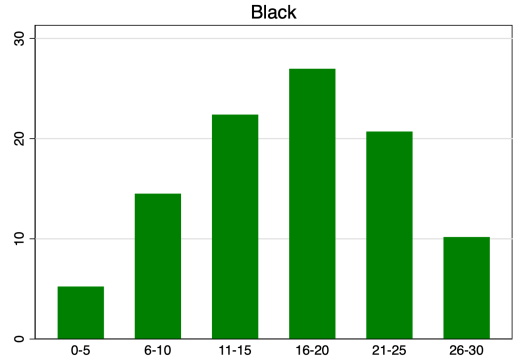
Result 4: In the Race-Belief Only treatment, the Asian-Black belief gap is significantly reduced but not eliminated.



Results: Recall

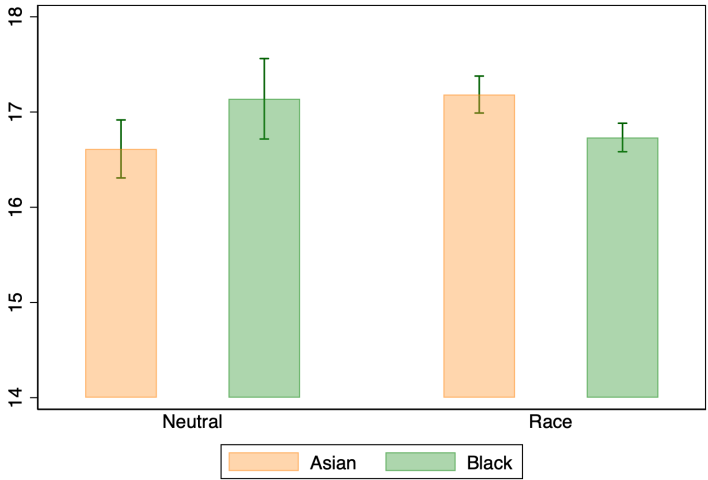
Recalled score distributions in Race Frame, Belief Only





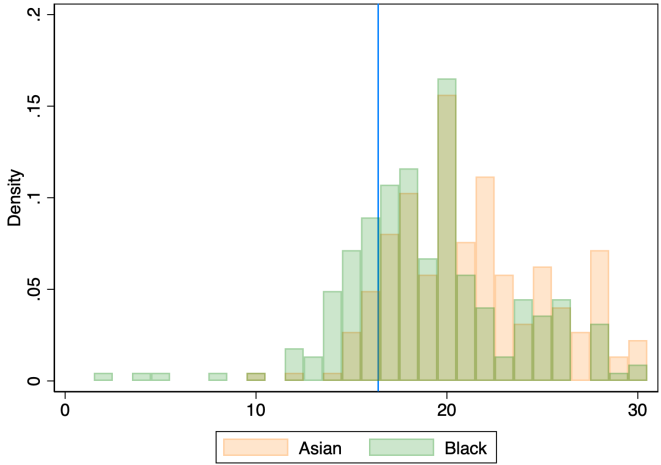
Results: Recall

Result 5: Means from recalled distributions are different for Asian and Black workers in the Race Frame.



Misuse of Statistics: Suggestive Evidence

Result 6: Employer beliefs for individual Asian workers are predominantly from the right tail of the group distribution, whereas there are more from the left tail for Black workers.



Taking Stock

- I use an experiment to show that starting with equal priors, employers who see race labels for workers exhibit significant discrimination against Black workers in hiring choices
- Employers hold a belief gap after making hiring choices in race-labeled treatment
- Taking away hiring reduces but doesn't eliminate belief gap
- Employers recall different distributions for Asian and Black workers in race labeled treatment

Contributions

- I contribute to the discrimination literature by providing a clear link between taste-based discrimination and statistical discrimination
- I identify a new source of biased beliefs that can lead to biased statistical discrimination
- Evidence on this belief distortion suggests that anti-discrimination policies that aim to reduce discrimination by providing group information may not be fully effective

Thank You

Christina Sun
University of California, Davis
Email: ucsun@ucdavis.edu