Diversity and Global Policy: Discrimination

Germain Gauthier

Bocconi University

We study the factors driving observed differences in group outcomes.

We focus on:

- **Socio-psychological factors**
  - Focus on gender differences in preferences
- **Culture, norms, and stereotypes**
  - Focus on their formation and persistence across time and space
- **(Unfair) institutions**
  - Focus on the history of slavery in the United States
- **Discrimination** (today's lecture)
  - Theoretical models of discrimination
  - Quasi-experimental evidence of discriminatory practices of employers

A Working Definition

Discrimination = treating someone differently based on characteristics such as gender, race, or religion

Obvious examples:

- Intentionally not considering certain employees for a promotion
- Refusing to rent your flat to people with an arabic-sounding name

Not-so-obvious examples:

- Consistently bringing up stereotypes in conversations
- Regularly excluding certain employees from social events

Measuring Discrimination is Hard

While documenting racial/gender disparities is relatively easy, identifying discrimination as the cause is more challenging.

Consider the following regression:

$$\log(\mathit{Wage})_i = \tau \mathit{Black}_i + X_i' \beta + \varepsilon_i$$

τ unlikely measures racial discrimination in the labor market because of omitted variable bias.

e.g., Black women are more likely to raise their kids alone, thus making it harder for them to put in the extra hours.

But can't we control for hours worked in X?

OK, but what about educational attainment, eloquence, social skills, etc.?

Ensuring that the researcher observes all that the decision-maker observes is a hopeless task.

Moreover, adding controls to a regression can obscure results.

e.g., if minority workers sort into industries where there is no or limited discrimination, then finding no racial gap in earnings after controlling for industry or employer fixed effects in a regression may indicate that there is no discrimination at the margin!

Experimental Approaches

Two common experimental approaches to overcome these challenges:

- **Audit studies:**
  - Send a pair of auditors matched on personal characteristics but different on some dimension (e.g., race) to apply for a real job.
- **Correspondence studies:**
  - Send fictitious resumes to real jobs that are perfectly similar albeit on one dimension (e.g., gender).

$\Rightarrow$ Both approaches have produced overwhelming evidence of labor market and rental market discrimination.

An Early Audit Study

Ayres and Siegelman (1995) sent pairs of testers to negotiate the purchase of a new automobile:

- Testers were chosen to have average attractiveness
- Testers in a pair bargained for the same model of car at the same dealership, usually within a few days of each other.
- All testers had the same bargaining script.
- Dealerships, testers, and the choice of which tester in the pair would be the first to enter the dealership were randomized.

Main result: White males are quoted lower prices than white women and blacks (men or women).

Limitations

- They require that both members of the auditor pair be identical in all dimensions that might affect productivity in employers' eyes (except for the trait that is being manipulated).
  - This is very unlikely!
- They are not double-blind: auditors know the purpose of the study.
  - This may generate conscious/subconscious motives among auditors to generate data consistent/inconsistent with their beliefs.
- They are expensive and typically have a small number of observations.
  - Statistical power is in the balance.

Correspondence Studies

Correspondence studies have been developed to address some of the more obvious weaknesses of the audit method.

Instead of sending actors, researchers create fictitious applicants in response to advertisements (*e.g., resumes, letters of interest, etc.*).

Pairs of fictitious applicants are in every way similar, except for the perceived minority trait (*e.g., a muslim-sounding name*).

Discrimination is estimated by comparing the outcomes for the fictitious applicants with and without the perceived minority trait.

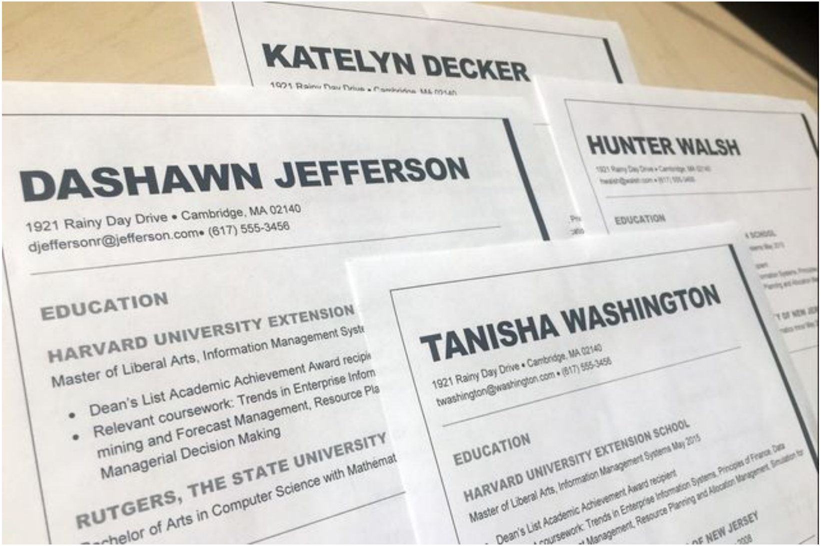

KATELYN DECKER
DASHAWN JEFFERSON
1921 Rainy Day Drive • Cambridge, MA 02140
djeffersonr@jefferson.com• (617) 555-3456
EDUCATION
HUNTER WALSH
EDUCATION
TANISHA WASHINGTON
1921 Rainy Day Drive • Cambridge, MA 02140
twashington@washington.com • (617) 555-3456
EDUCATION
HARVARD UNIVERSITY EXTENSION SCHOOL

Table 1 Labor market correspondence studies

Paper	Country	CVs / apps	Vacancies	Effect (call-back ratio)	Theory
Galarza and Yamada (2014) Trait: Ethnicity; attractiveness	Peru	4820	1205	White-to-indigenous ratio: 1.8 Low attractiveness hurts white females	No
Eriksson and Rooth (2014) Trait: Unemployment duration	Sweden	8466	-	Employed to long-term unemployed: 1.25	No
Blommaert et al. (2014) Trait: Arabic name	Netherlands	636	-	Dutch-to-foreign: 1.62 (unconditional ratio). No difference, if views held fixed	No
Nunley et al. (2014) Trait: Race	United States	9396	-	White-to-black: 1.18 (unconditional)	Inconsistent with statistical discrimination, consistent with taste-based discrimination
Ghayad (2013) Trait: Unemployment duration	United States	3360	600	Employed-to-unemployed: 1.47	No
Bartoš et al. (2013) Trait: Ethnicity (Roma, Asian, Turkish)	Czech Republic and Germany	274 (Czech R.) 745 (Ger.)	-	Czech-to-Vietnamese: 1.34 Lower requests for CVs if candidate is Turkish	Consistent with attention discrimination
Wright et al. (2013) Trait: Religion/ethnicity	United States	6400	1600	White-to-Muslim: 1.58	Consistent with theoretical models of secularization and cultural distate theory
Kroft et al. (2013) Trait: Unemployment duration	United States (largest 100 MSAs)	12,054	3040	1 log point change in unemployment duration: 4.7 percentage points lower call-back probability	No

Continued

Table 1 Labor market correspondence studies—cont'd

Paper	Country	CVs / apps	Vacancies	Effect (call-back ratio)	Theory
Baert et al. (2013) Trait: Nationality (Turkish-sounding name)	Belgium	752	376	Flemish-to-Turkish: 1.03 to 2.05, depending on the occupation	No
Bailey et al. (2013) Trait: Sexual orientation	United States	4608	1536	No effect	No
Ahmed et al. (2013) Trait: Sexual orientation	Sweden	3990	-	Heterosexual-to-homosexual (male): 1.14 Heterosexual-to-homosexual (female): 1.22	No
Acquisti and Fong (2013) Traits:Sexual orientation and religion	United States	4183	-	Christian-to-Muslim: 1.16	No
Patacchini et al. (2012) Traits: Sexual orientation and attractiveness	Italy	2320	-	Heterosexual-to-Homosexual: 1.38	No
Kaas and Manger (2012) Trait: Immigrant (race/ ethnicity)	Germany	1056	528	German-to-Turkish: 1.29 (if no reference letter is included)	Consistent with statistical discrimination
Maurer-Fazio (2012) Trait: Ethnicity	China	21,592	10,796	Han-to-Mongolian: 1.36 Han-to-Tibetan: 2.21	No
Jacquemet and Yannelis (2012) Trait: Race/nationality	United States	330	990	English-to-foreign names: 1.41 English-to-Black names: 1.46	Consistent with patterns of ethnic homophily
Ahmed et al. (2012) Trait: Age	Sweden	466	-	31-year old to 46-year old: 3.23	No
Oreopoulos (2011) Trait: Nationality (and race)	Canada	12,910	3225	English name-to-Immigrant: ranged from 1.39 to 2.71 (against Indian Pakistani and Chinese applicants)	No
Carlsson (2011) Trait: Gender	Sweden	3228	1614	Female-to-male: 1.07	No
Booth et al. (2011) Trait: Ethnicity	Australia	Above 4000	-	White-to-Italian: 1.12 White-to-Chinese: 1.68	No

Booth and Leigh (2010) Trait: Gender	Australia	3365	-	Female-to-male: 1.28 (female-dominated professions)	No
Riach and Rich (2010) Trait: Age	United Kingdom	1000+	-	2.64 favoring younger candidates	No
Rooth (2009) Trait: Attractiveness/obesity	Sweden	1970	985	Nonobese/attractive-to-obese/unattractive: ranged from 1.21 to 1.25 (but higher for some occupations)	No
McGinnity et al. (2009) Trait: Nationality/race	Ireland	480	240	1.8, 2.07, 2.44 in favor of Irish and against Asians, Germans, and Africans, respectively	No
Banerjee et al. (2009) Traits: Caste and religion	India	3160	371	Upper caste-to-other backward castes: 1.08 (software jobs, insignificant), 1.6 (call-center jobs)	No
Lahey (2008) Trait: Age	United States	App. 4000	-	Young-to-older: 1.42	No
Petit (2007) Traits: Age, gender, number of children	France	942	157	Ranged from 1.13 to 2.43 against 25-year old, childless women	No
Bursell (2007) Trait: Ethnicity	Sweden	3552	1776	Swedish-to-foreign names: 1.82	Inconsistent with statistical discrimination
Bertrand and Mullainathan (2004) Trait: Race	United States	4870	1300+	White-to-African-American: 1.5 (1.22 for females in sales jobs)	No
Jolson (1974) Trait: Race and religion	United States	300	-	White-to-black: 4.2 for selling positions	No

MSAs; Metropolitan Statistical Areas.

Table 2 Rental market papers

Study	Country	Inquiries	Effect	Theory
Carlsson and Eriksson (2014) Trait: Minority status (Arabic name)	Sweden	5827	Swedish-to-Arabic (females): 1.37 Swedish-to-Arabic (males): 1.62	No
Ewens et al. (2014) Trait: Race	United States	14,237	White-to-black: 1.19	Consistent with statistical discrimination, inconsistent with taste-based discrimination
Bartoš et al. (2013) Trait: Minority status (Roma or Asian name)	Czech Republic and Germany	1800	Czech-to-minority: 1.27 (site available), 1.9 (pooled Asian and Roma names)	Consistent with attention discrimination
Hanson and Hawley (2011) Trait: Race	United States	9456	White-to-African-American: 1.12 (varied by neighborhood and unit type)	Consistent with statistical discrimination
Baldini and Federici (2011) Trait: Immigrant status; Language ability	Italy	3676	Italian-to-East European: 1.24 Italian-to-Arab: 1.48	No
Ahmed et al. (2010) Trait: Minority status (Arabic name)	Sweden	1032	Swedish-to-Arab/Muslim: 1.44 (no information), 1.24 (detailed information about the applicant)	No
Bosch et al. (2010) Trait: Immigrant status	Spain	1809	Spanish-to-Moroccan: 1.44 (no information), 1.19 (with positive information)	No
Ahmed and Hammarstedt (2009) Trait: Sexual orientation	Sweden	408	Straight-to-gay: 1.27	No
Ahmed and Hammarstedt (2008) Trait: Immigrant (race/ ethnicity/religion)	Sweden	1500	Swedish-to-Arab male: 2.17	No
Carpusor and Loges (2006) Trait: Race/ethnicity (Arab, African-American)	United States (Los Angeles County)	1115	White-to-Arab: 1.35 White-to-black: 1.59, conditional on hearing back, 1.98 unconditional	No

Limitations

Correspondence studies address some key weaknesses of the audit methodology.

But they share other weaknesses with audit studies and have some unique limitations of their own:

- Correspondence approaches cannot be taken to **future stages**.
  - *e.g., the interview stage, job-offer stage, wage-setting stage, lease-signing stage, etc.*
- Both methods raise **ethical concerns**.
  - This is wasted time for employers.
  - It may also shift their prior beliefs in damaging ways.

Very few empirical studies attempt to relate their results to theories of discrimination... But what are these theories to begin with?

Why do people discriminate?

Even when disparities can be attributed to discrimination, the causes of discriminatory behavior may differ.

Three common theories:

- **Taste-based discrimination**
  - Based on prejudice (or “preferences” in economic terms)
- **Statistical discrimination**
  - Based on valid statistical inference in contexts with limited information
- **Systemic discrimination**
  - A broader perspective that moves beyond direct discrimination and focuses on barriers encountered across the life cycle

Taste-based Discrimination

Wages differ between groups because prejudiced employers have a willingness to pay to avoid hiring minority workers.

This is the theory of taste-based discrimination (Becker, 2010).

Taste-based Discrimination

Two groups: a is majority, b is minority

They are perfect substitutes in production (i.e., same productivity).

Some employers are prejudiced against group b

d: “coefficient of discrimination” (i.e., firm “taste” parameter)

Employers maximize utility (not profits):

$$U = pF(N_b + N_a) - w_a N_a - w_b N_b - dN_b,$$

where p is price, F is production function, N_g is employment of members of group g, and w_g is wage paid to members of group g.

Employers

Prejudiced employers have $d \geq 0$.

For them, the price of hiring a worker from group b is $w_b + d$.

Employer i will hire workers from group b only if $w_a \geq d_i + w_b$.

Let $G(d)$ denote the CDF of prejudice parameter d among firms.

Hiring

Firms choose N_a and N_b according to:

$$\frac{dU}{dN_a} = 0 \;\; \Rightarrow \;\; pF'(N_b + N_a) = w_a$$

for firms that hire workers from group a and

$$\frac{dU}{dN_b} = 0 \;\; \Rightarrow \;\; pF'(N_b + N_a) = w_b + d$$

for firms that hire workers from group b.

$\Rightarrow$ Firm i hires b workers if $w_b + d_i \leq w_a$.

Equilibrium

Treat price p as fixed

Market demand: $N_b^d(w_a, w_b; G(d))$, $N_a^d(w_a, w_b; G(d))$

Market supply: $N_a^s(w_a)$, $N_b^s(w_b)$

Wages are determined when supply and demand intersect:

$$N_a^d(w_a, w_b; G(d)) = N_a^s(w_a)$$

$$N_b^d(w_a, w_b; G(d)) = N_b^s(w_b)$$

Some Intuition

Two key insights:

- **Sorting**
  - b workers are employed by the least prejudiced firms (i.e., firms i for which $w_b + d_i \leq w_a$)
  - Only marginal firms hire both groups (i.e., the firms m for which $w_b + d_m = w_a$)
- **Marginal preferences**
  - The wage gap is determined by the preferences of the least prejudiced employer who hires b workers (i.e., d_m), not by average prejudice

Graphical Illustration

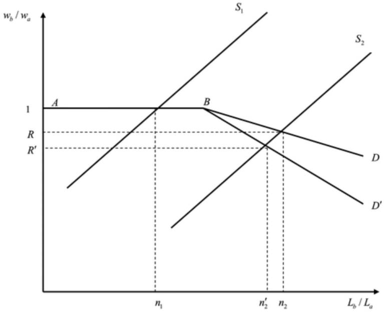


Fig. 2.— Relationship between racial tastes and the relative wages and relative supply of blacks and whites. The figure shows how the equilibrium ratio of black to white wages responds to three sets of market conditions. When the relative supply of black workers is small relative to the number of unprejudiced employers, as is the case when supply is as depicted by S_1, the marginal discriminator is unprejudiced and there is no racial wage gap in equilibrium. When the distribution of racial preferences among employers is held constant, a shift out in the relative supply of black workers (from S_1 to S_2) requires that more prejudiced employers hire blacks, and the ratio of black to white wages falls from one to R. When the relative supply of black workers is held constant, an increase in prejudice among employers likely to be the marginal discriminator (which causes the relative demand curve to rotate from ABD to ABD'), further reduces the equilibrium ratio of black to white wages to R'.

Limitations and Counter-intuitive Predictions

Strange feature (Becker, Arrow): prejudiced employers will be **driven out** of the market in a **long-run competitive setting** because they earn lower profits than other firms...

In other words, the seminal model doesn't predict long-run equilibrium wage differentials!

Arrow: Becker's employer discrimination model "predicts the absence of the phenomenon it was designed to explain"!

In practice, adding market frictions (*e.g., search and adjustment costs*) is enough to generate equilibrium wage differentials.

Empirical Evidence

Most people do not admit or may not recognize that they are discriminating, let alone attribute it to prejudice...

However, according to this model, the racial prejudice of the marginal employer of black employees determines the racial wage gap.

Charles and Guryan (2008) estimate the 10^{th}, 50^{th}, and 90^{th} percentile of racial prejudice in each US state.

Consistent with taste-based discrimination, they find that the 10^{th} percentile of racial prejudice best predicts the racial wage gap.

Statistical Discrimination

What if employers have nothing against minorities *per se* but simply make valid statistical inferences in contexts with limited information?

This is the theory of statistical discrimination (Aigner and Cain, 1977).

Setup

Employers base hiring decisions on an indicator of skill y (say, a test) that measures a worker's true skill level q:

$y = q + u$, where $u \sim N(0, \sigma_u^2)$ and $q \sim N(\alpha, \sigma_q^2)$.

y is a linear combination of two random normal variables, so we have:

$$y \sim N(\alpha, \sigma_q^2 + \sigma_u^2)$$

Employers observe y but not q.

They use y to extract information about q: $\hat{q} = \mathbb{E}[q|y]$.

After some calculus, we can show that:

$$\hat{q} = (1 - \gamma)\alpha + \gamma y$$

This is a signal extraction problem!

In expectation, a worker's productivity is a weighted average of her test score y and the group average α, where weights are determined by γ:

$$\gamma = \frac{var(q)}{var(q) + var(u)}$$

γ is smaller if the test is less informative (higher $var(u)$). Then employers put more weight on group average α to form their expectations.

Now consider two groups of workers: whites and blacks, possibly different means (α_w and α_b) and variances of q and u.

Employers pay workers based on data available for each group:

$$\hat{q}_w = (1 - \gamma_w)\alpha_w + \gamma_w y$$

$$\hat{q}_b = (1 - \gamma_b)\alpha_b + \gamma_b y$$

In general, $\gamma_w \neq \gamma_b$ if variances of q, u differ.

If the test is more informative for whites ($var(u_b) \geq var(u_w)$), then $\gamma_w \geq \gamma_b$, and employers put more weight on individual test scores for whites relative to blacks.

Let's build some intuition by considering two special cases:

- Mean differences, but equal variances

$$\alpha_b \leq \alpha_w, var(u_w) = var(u_b) \text{ , and } var(q_w) = var(q_b)$$

- Equal means, but different variances

$$\alpha_b = \alpha_w, var(u_w) \geq var(u_b) \text{ , and } var(q_w) = var(q_b)$$

Mean Differences, Equal Variances

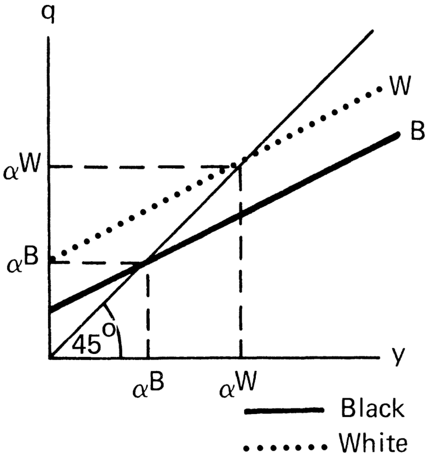


Figure 5. Prediction of Productivity (q), by Race and Test Score (y), Assuming the Slopes Are Equal.

Equal Means, Different Variances

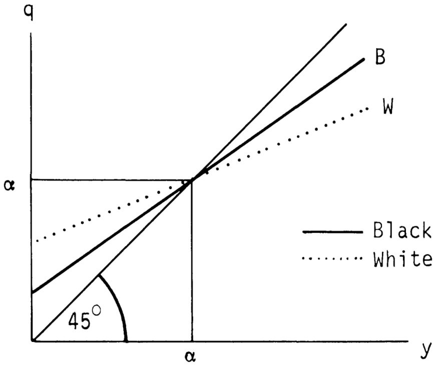


Figure 1A. Predictions of Productivity (q) by Race and Test Score (y), Assuming a Steeper Slope for Blacks.

Empirical Evidence

There is strong evidence of statistical discrimination in many settings.

e.g., On the market for sports cards, see List (2004); on the commercial sex market in Singapore, see Li et al. (2018).

The theory of statistical discrimination suggests that providing information about characteristics correlated with race can reduce discrimination.

e.g., Wozniak (2015) finds that drug testing increased the employment of blacks.

e.g., Doleac and Hansen (2020) find that banning the box reduced the employment of low-skill young black men by 3.4 percentage points.

Quiz

Is statistical discrimination a justification for discriminatory practices?

Systemic Discrimination

Critique of sociologists: Both standard workhorse economics models of discrimination fail to understand discrimination as a **systemic problem**.

This problem takes many forms.

For one, discrimination is experienced by individuals **at every step of their lives**.

Observed group differences result from the **cumulative effects** of discrimination across individuals' lives.

e.g., born in a segregated neighbourhood → not pushed to perform well in school → mediocre career choice → limited opportunities for career promotions → etc.

A Broader Perspective on Discrimination

- Organizations can discriminate irrespective of the intentions of their members:
  - *e.g., Filling in vacancies via referral networks*
- Historic discrimination has contemporary consequences:
  - *cf. Lecture on Slavery and Segregation in the US*
- Ostensibly minor forms of discrimination can have major consequences:
  - "Micro-agressions" have been shown to affect mental health and physical well-being.
  - *e.g., Being repeatedly followed by a security guard at a store, repeatedly confronted with racial slights at work, etc.*
- Perceived discrimination also matters for mental health, depression, stress, and related health outcomes.

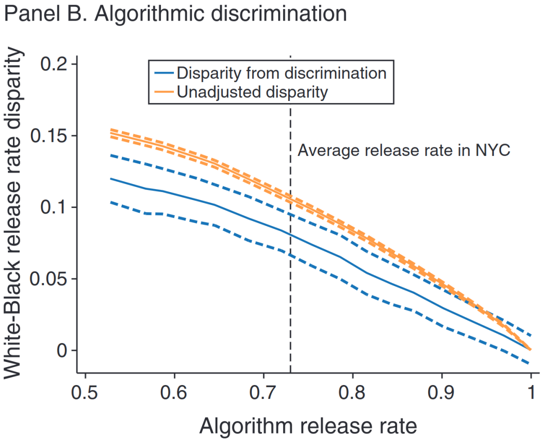
Modern Issues: Algorithmic Bias

In the digital age, many algorithms are developed to help humans make decisions (or make them in their place).

The choice of the data, the model, and its outcomes may all bake in biased algorithmic decisions:

- **Data:** Many datasets are historical records of human decisions and may be potentially biased.
- **Models:** Depending on the model's complexity, understanding *why* the machine makes this decision can be challenging (often referred to as the "black box" problem).
- **Outcomes:** Often, models predict proxies of outcomes of interest. This may seriously bias algorithmic decisions.

Bail Decisions



Source: Arnold et al. (2021)

Health Decisions

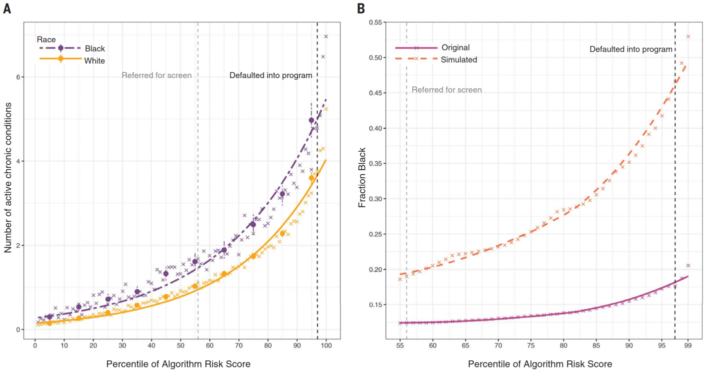


Fig. 1. Number of chronic illnesses versus algorithm-predicted risk, by race. (**A**) Mean number of chronic conditions by race, plotted against algorithm risk score. (**B**) Fraction of Black patients at or above a given risk score for the original algorithm ("original") and for a simulated scenario that removes algorithmic bias ("simulated": at each threshold of risk, defined at a given percentile on the *x* axis, healthier Whites above the threshold are replaced with less healthy Blacks below the threshold, until the marginal patient is equally healthy). The × symbols show risk percentiles by race; circles show risk deciles with 95% confidence intervals clustered by patient. The dashed vertical lines show the auto-identification threshold (the black line, which denotes the 97th percentile) and the screening threshold (the gray line, which denotes the 55th percentile).

Source: Obermeyer et al. (2019)

ChatGPT

topic	high probability words	all GPT-3	matched GPT-3
life	really, time, want, going, sure, lot, feel, little, life, things	0.018	0.010
family	baby, little, sister, child, girl, want, children, father, mom, mama	0.014	0.007
appearance	woman, girl, black, hair, white, women, looked, look, face, eyes	0.007	0.006
politics	people, country, government, president, war, american, world, chinese, political, united states	-0.008	-0.003
war	men, war, soldiers, soldier, general, enemy, camp, fight, battle, fighting	-0.008	-0.006
machines	plane, time, air, ship, machine, pilot, space, computer, screen, control	-0.008	-0.004

Table 1: Feminine and masculine main characters are associated with different topics, even in the matched prompt setup. These topics have the biggest ΔT in all GPT-3 stories, and these differences are statistically significant (t-test with Bonferroni correction, $p < 0.05$).

Source: Lucy and Bamman (2021)

Conclusion

Discrimination is often the “usual suspect” when mentioning group differences such as wage gaps.

And for good reason because there is considerable quasi-experimental evidence of discrimination in many markets.

Understanding the exact channels through which discrimination operates remains a source of disagreement among social scientists.

Recently, the rise of algorithms for decision-making has raised many issues and become an active area of research.

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Ilr Review*, 30(2):175–187.

Arnold, D., Dobbie, W., and Hull, P. (2021). Measuring racial discrimination in algorithms. In *AEA Papers and Proceedings*, volume 111, pages 49–54. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Ayres, I. and Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review*, pages 304–321.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Charles, K. K. and Guryan, J. (2008). Prejudice and wages: an empirical assessment of becker's the economics of discrimination. *Journal of political economy*, 116(5):773–809.

Doleac, J. L. and Hansen, B. (2020). The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden. *Journal of Labor Economics*, 38(2):321–374.

Li, H., Lang, K., and Leong, K. (2018). Does competition eliminate discrimination? evidence from the commercial sex market in singapore. *The Economic Journal*, 128(611):1570–1608.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics*, 119(1):49–89.

Lucy, L. and Bamman, D. (2021). Gender and representation bias in gpt-3 generated stories. In *Proceedings of the Third Workshop on Narrative Understanding*, pages 48–55.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453.

Wozniak, A. (2015). Discrimination and the effects of drug testing on black employment. *Review of Economics and Statistics*, 97(3):548–566.

Tips for the Exam

Understand the limitations of correlational evidence of discrimination.

Know the various quasi-experimental approaches to measure discrimination in the field.

Know the three theoretical perspectives on discrimination.

Be aware that discrimination is not typically human – algorithms can also be biased and should be carefully designed and audited.