Discrimination

Miguel A. Fonseca

University of Exeter

Economics of Discrimination

Discrimination is a central topic in today's society

- gender pay gap
- lack of diversity in boards
- under-representation of minorities

Why is that the case?

Modeling Discrimination in Economics

There are several theoretical approaches to modeling discrimination in the labour market (or in other contexts).

1. Taste-based discrimination
   - Firms make have intrinsic preferences to hire workers matching a particular ethnic/gender profile (Becker, 1972)
   - Homophily (Currarrini, Jackson and Pin, 2008)

Modeling Discrimination in Economics

2. Statistical discrimination
   - Using observable characteristics to make statistical inference about productivity (Arrow, 1972; Phelps, 1972)
   - The precision of the signal employers get may be lower for the discriminated-against group (Altonji and Blank, 1999)

Testing Discrimination in Economics

The first type is undesirable, illegal, and economically inefficient.

- The challenge is for labour economists to distinguish it from the second type.

We will explore different types of experiments that tackle this issue:

- Written Contact Field Experiments
- Personal Contact/Audit Field Experiments

Written Contact Field Experiments

Written Contact experiments involve sending carefully-matched sets of written job applications/CVs in response to advertised vacancies.

- Some studies send *unsolicited* job applications to test for preferential treatment in employer responses, rather than biases in market outcomes.

Written Contact Field Experiments

In order to avoid detection, letters are not identical, but in all essential aspects (e.g. qualifications, experience), the candidates are equivalent.

- In this way, the only distinguishing characteristic is gender, ethnicity, and/or disability

Letter/CV types are crossed across the relevant types to ensure the style of application does not bias response rates.

Written Contact Field Experiments

Advantages:

- Great degree of control over experimental manipulation.
- No scope for demand effects.
- Low cost ($=$ large sample size $=$ good statistical power $\rightarrow$ more treatment variables in one study).
- Even easier now with websites like `monster.com` or `jobs.theguardian.com`.

Written Contact Field Experiments

Disadvantages:

- Crude outcome measure: a callback for an interview does not mean the job.
- We cannot obtain salary data, an important labour market outcome.
- Some characteristics not easy to convey through name (e.g. ethnicity).
- The studies only focus on one labour market mechanism. If one group relies more heavily on social networks to obtain employment, this will bias results.

Bertrand and Mullainathan (2004)

Test whether there is gender and race discrimination in Boston and Chicago labor markets.

- They focus on sales, admin support, clerical services and customer services jobs

They create a bank of resumes, using as templates actual resumes posted in websites

- This ensures applications are realistic
- Names and identifying information are changed to safeguard original job applicants' privacy
- They only use resumes posted ≥ 6 months before the start of the experiment.

Bertrand and Mullainathan (2004)

They classify the resumes within each job category as high or low quality based on:

- Labor mkt experience, career profile, gaps in employment, other skills

The authors add to high-quality resumes other features like:

- Language skills, volunteering, extra computer skills, military service, email
- To avoid over-qualification, they add these skills at random

Importantly, this is done **before** gender and race assignments are done.

Bertrand and Mullainathan (2004)

B & M then generated identities for the fictitious applicants

- Name, phone number, address, email

To select which names are uniquely African-American/White, they used frequency data from birth certificates registered in Massachusetts between 1974 and 1979.

- To check for distinctiveness they ran a survey where respondents had to associate several features of a person with a given name.
- Names who were associated with the two ethnicities got picked.

Applicants in each race/sex/city/resume quality cell are allocated the same phone number

- This is done in order to track callback rates.
- Phone numbers didn't work – calls went straight to voicemail.

Bertrand and Mullainathan (2004)

The experiment was carried out between July 2001 and January 2002 in Boston, and between July 2001 and May 2002 in Chicago.

- This allows the authors to compare tight and slack labour market periods.

Over the period, B&M surveyed job ads in *The Boston Globe* and *The Chicago Tribune*.

- They did not consider ads requesting the applicant to call or apply in person.
- Most ads asked applicants to fax or mail their application.
- They logged all available information about the employer.

Bertrand and Mullainathan (2004)

For each job B&M sampled four resumes (two high quality & two low quality) that fit job ad.

- One resume of each quality is selected at random to get African-American names
- The other resume of similar quality got White name.

B&M used both gendered names for sales jobs, but only female names for admin or clerical jobs

- Callback rates were low for male applicants in the latter two categories.

B&M responded to more than 1,300 ads and sent out almost 5,000 resumes.

Bertrand and Mullainathan (2004)

Response measure: phone or email callback for an interview.

- They cannot verify snail mail callbacks because addresses were fake.
- Snail mail is rarely used for this purpose, so potential bias is low.

Bertrand and Mullainathan (2004)

TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (*p*-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

Notes: The table reports, for the entire sample and different subsamples of sent resumes, the callback rates for applicants with a White-sounding name (column 1) an an African-American-sounding name (column 2), as well as the ratio (column 3) and difference (column 4) of these callback rates. In brackets in each cell is the number of resumes sent in that cell. Column 4 also reports the *p*-value for a test of proportion testing the null hypothesis that the callback rates are equal across racial groups.

Bertrand and Mullainathan (2004)

TABLE 2—DISTRIBUTION OF CALLBACKS BY EMPLOYMENT AD

Equal Treatment:	No Callback	1W + 1B	2W + 2B
88.13 percent	83.37	3.48	1.28
[1,166]	[1,103]	[46]	[17]
Whites Favored (WF):	1W + 0B	2W + 0B	2W + 1B
8.39 percent	5.59	1.44	1.36
[111]	[74]	[19]	[18]
African-Americans Favored (BF):	1B + 0W	2B + 0W	2B + 1W
3.48 percent	2.49	0.45	0.53
[46]	[33]	[6]	[7]
Ho: *WF* = *BF*			
$p = 0.0000$			

Bertrand and Mullainathan (2004)

TABLE 4—AVERAGE CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES AND RESUME QUALITY

	Low	High	Ratio	Difference (*p*-value)
Panel A: Subjective Measure of Quality (Percent Callback)				
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)
Panel B: Predicted Measure of Quality (Percent Callback)				
	Low	High	Ratio	Difference (*p*- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

Notes: Panel A reports the mean callback percents for applicant with a White name (row 1) and African-American name (row 2) depending on whether the resume was subjectively qualified as a lower quality or higher quality. In brackets is the number of resumes sent for each race/quality group. The last column reports the *p*-value of a test of proportion testing the null hypothesis that the callback rates are equal across quality groups within each racial group. For Panel B, we use a third of the sample to estimate a probit regression of the callback dummy on the set of resume characteristics as displayed in Table 3. We further control for a sex dummy, a city dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). We then use the estimated coefficients on the set of resume characteristics to estimate a predicted callback for the remaining resumes (two-thirds of the sample). We call "high-quality" resumes the resumes that rank above the median predicted callback and "low-quality" resumes the resumes that rank below the median predicted callback. In brackets is the number of resumes sent for each race/quality group. The last column reports the *p*-value of a test of proportion testing the null hypothesis that the callback percents are equal across quality groups within each racial group.

Bertrand and Mullainathan (2004)

TABLE 5—EFFECT OF RESUME CHARACTERISTICS ON LIKELIHOOD OF CALLBACK

Dependent Variable: Callback Dummy Sample:	All resumes	White names	African-American names
Years of experience (*10)	0.07	0.13	0.02
	(0.03)	(0.04)	(0.03)
Years of experience^2 (*100)	−0.02	−0.04	−0.00
	(0.01)	(0.01)	(0.01)
Volunteering? (Y = 1)	−0.01	−0.01	0.01
	(0.01)	(0.01)	(0.01)
Military experience? (Y = 1)	−0.00	0.02	−0.01
	(0.01)	(0.03)	(0.02)
E-mail? (Y = 1)	0.02	0.03	−0.00
	(0.01)	(0.01)	(0.01)
Employment holes? (Y = 1)	0.02	0.03	0.01
	(0.01)	(0.02)	(0.01)
Work in school? (Y = 1)	0.01	0.02	−0.00
	(0.01)	(0.01)	(0.01)
Honors? (Y = 1)	0.05	0.06	0.03
	(0.02)	(0.03)	(0.02)
Computer skills? (Y = 1)	−0.02	−0.04	−0.00
	(0.01)	(0.02)	(0.01)
Special skills? (Y = 1)	0.05	0.06	0.04
	(0.01)	(0.02)	(0.01)
Ho: Resume characteristics effects are all zero (*p*-value)	54.50 (0.0000)	57.59 (0.0000)	23.85 (0.0080)
Standard deviation of predicted callback	0.047	0.062	0.037
Sample size	4,870	2,435	2,435

Notes: Each column gives the results of a probit regression where the dependent variable is the callback dummy. Reported in the table are estimated marginal changes in probability for the continuous variables and estimated discrete changes for the dummy variables. Also included in each regression are a city dummy, a sex dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). Sample in column 1 is the entire set of sent resumes; sample in column 2 is the set of resumes with White names; sample in column 3 is the set of resumes with African-American names. Standard errors are corrected for clustering of the observations at the employment-ad level. Reported in the second to last row are the *p*-values for a χ^2 testing that the effects on the resume characteristics are all zero. Reported in the second to last row is the standard deviation of the predicted callback rate.

Bertrand and Mullainathan (2004)

TABLE 7—EFFECT OF JOB REQUIREMENT AND EMPLOYER CHARACTERISTICS ON RACIAL DIFFERENCES IN CALLBACKS

Job requirement:	Sample mean (standard deviation)	Marginal effect on callbacks for African-American names
Any requirement? (Y = 1)	0.79	0.023
	(0.41)	(0.015)
Experience? (Y = 1)	0.44	0.011
	(0.49)	(0.013)
Computer skills? (Y = 1)	0.44	0.000
	(0.50)	(0.013)
Communication skills? (Y = 1)	0.12	−0.000
	(0.33)	(0.015)
Organization skills? (Y = 1)	0.07	0.028
	(0.26)	(0.029)
Education? (Y = 1)	0.11	−0.031
	(0.31)	(0.017)
Total number of requirements	1.18	0.002
	(0.93)	(0.006)

Employer characteristic:	Sample mean (standard deviation)	Marginal effect on callbacks for African-American names
Equal opportunity employer? (Y = 1)	0.29	−0.013
	(0.45)	(0.012)
Federal contractor? (Y = 1)	0.11	−0.035
(N = 3,102)	(0.32)	(0.016)
Log(employment)	5.74	−0.001
(N = 1,690)	(1.74)	(0.005)
Ownership status:		
(N = 2,878)		
Privately held	0.74	0.011
		(0.019)
Publicly traded	0.15	−0.025
		(0.015)
Not-for-profit	0.11	0.025
		(0.042)
Fraction African-Americans in employer's zip code	0.08	0.117
(N = 1,918)	(0.15)	(0.062)

Bertrand and Mullainathan (2004)

How do theories of discrimination deal with B&M's results?

- The lower returns to credentials for African-Americans
- The relative uniformity of the race gap across occupations

Taste-based models where the prejudice comes from co-workers and/or customers don't do well

- Not much variation in the racial gap wrt occupation and industry, particularly customer facing jobs

Employer racial prejudice could explain some of the results, but it does not explain the fact that African-Americans get lower returns to credentials.

- As a candidate's skills increase, it becomes more "expensive" to turn down a qualified candidate

Bertrand and Mullainathan (2004)

Statistical discrimination models based on using race to proxy unobservable skills struggle to explain the differential responsiveness to credentials.

- In fact, they should predict the *opposite* effect, particularly verifiable accreditations

SD models based on "signal precision" would argue the same signal is more informative for Whites than African-Americans

- In these models, African-Americans receive a lower return to a given skill because employers put less weight on that skill
- But in this experiment, the signal is the same for both groups
- And the signal is verifiable (e.g. university degree)

Bertrand and Mullainathan (2004)

B&M argue stereotypes or lexicographic decision-making, may play a role in employer decision-making.

In a later paper, Bertrand, Chugh and Mullainathan argue that implicit biases may be at play.

Bartos et al. (2016) put forward a very different explanation, based on cognitive/time constraints and statistical discrimination.

Bartos et al. (2016)

Bartos et al propose that individuals have limited attention to the information that is available to them.

This is crucial in the way most selection processes work. The Economist (2012) describes how HR process applications as:

They [human resource staff] look at a CV for ten seconds and then decide whether or not to continue reading. If they do, they read for another 20 seconds, before deciding again whether to press on, until there is either enough interest to justify an interview or to toss you into the 'no' pile.

A Model of Attention Discrimination

There are two stages to the decision-making process:

1. the DM first observes the applicant's group of ethnic origin G
   - the DM decides whether to pay additional attention to the applicant and invite her to an interview
2. the DM then obtains more information about the applicant

The role of stage 1 is to preselect applicants.

A Model of Attention Discrimination

For the DM, the applicant is of an unknown payoff π.

$$\pi = q - d_G \tag{1}$$

q is an unknown objective quality of the applicant (e.g. skill);

- assume it is normally distributed with known parameters

d_G is the DM's known distaste toward group G

A Model of Attention Discrimination

Further assume that quality can be decomposed in three parts:

$$q = q_G + q_1 + q_2 \tag{2}$$

q_G is the average quality of group G (observable)

q_1 is quality information that can be gleaned from CV

q_2 is quality information that can only be gotten from interview

A Model of Attention Discrimination

The DM earns π if it accepts the applicant and R if it rejects the applicant, but always pays inspection costs.

When all costs of information are sunk, the applicant is accepted if and only if $q - d_G > R$

DEFINITION (The DM's first-stage problem): *Upon observing G, the DM first chooses whether to incur C_1 and receive additional information, or to reject or invite the applicant without it. He chooses the action that maximizes the expected payoff:*

$$\begin{aligned} payoff(reject) &= R \\ payoff(invite) &= E\big[\max(R, q - d_G)\big] - C_2 \\ payoff(info) &= E\Big[\max\Big(R, E\big[\max(R, q - d_G)\,|\,q_1\big] - C_2\Big)\Big] - C_1. \end{aligned}$$

A Model of Attention Discrimination: Cherry-Picking vs. Lemon Dropping Markets

There are two types of markets that merit discussion

A cherry-picking market is a selective market:

- based only on group attributes, *payoff(reject)* $>$ *payoff(invite)*
- e.g. a labour market with many applicants for few posts, but few fit candidates

A lemon-dropping market is a non-selective market:

- based only on group attributes, *payoff(reject)* $\leq$ *payoff(invite)*
- e.g. a housing rental market where average applicant is acceptable

A Model of Attention Discrimination: Cherry-Picking vs. Lemon Dropping Markets

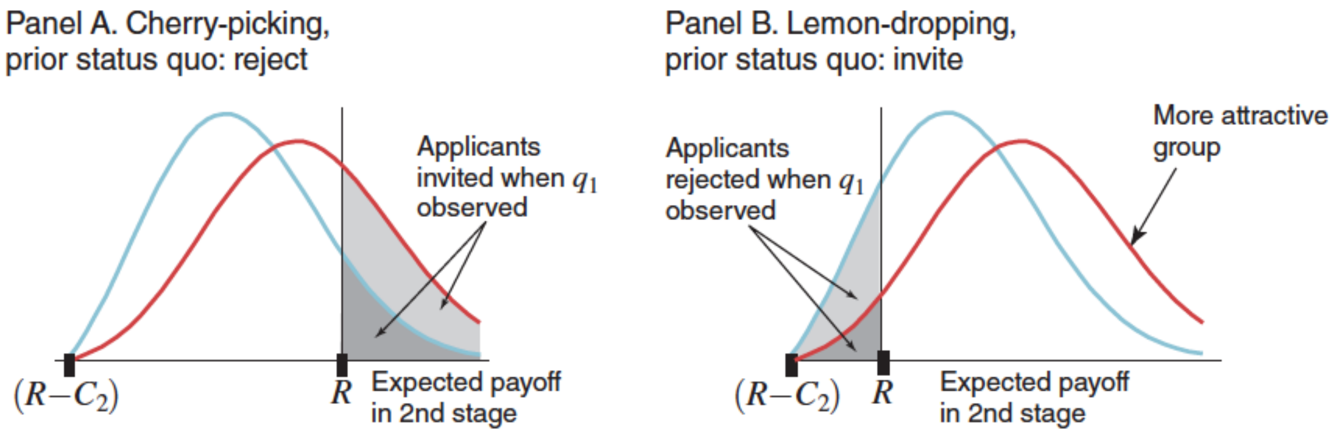


FIGURE 1. EXPECTED BENEFITS FROM INFORMATION ACQUISITION IN THE FIRST STAGE

A Model of Attention Discrimination: Cherry-Picking vs. Lemon Dropping Markets

Taste-based discrimination shifts a distribution to the left

- There are fewer candidates over the threshold R

Statistical discrimination works through the mean and variance of the distribution itself

- Lower mean and/or higher variance, more statistical discrimination.

A Model of Attention Discrimination: Cherry-Picking vs. Lemon Dropping Markets

The model makes two important predictions:

1. Endogenous attention disadvantages a dissimilar group to the DM in the cherry-picking market and helps it in the lemon-dropping market.

2. If both groups are either in the cherry-picking market or both are in the lemon-dropping market, then the probability that an applicant from a less attractive group is accepted is (weakly) lower if he is known to be from G a priori rather than when he is first considered to be from a general population and his membership in G is revealed only before the final selection decision.

Testing Attention Discrimination: A Lemon Dropping Market

Bartos et al. (2016) study ethnic discrimination in the Czech rental housing market.

- Roma and Vietnamese communities

Both groups are economically and socially disadvantaged in the Czech Republic

- The unemployment rate of Roma is estimated to be 38% (9.4% nationally)
- High school graduation rates are 47% and 33% for Vietnamese and Roma (84% for ethnic majority)
- Poll data revealed that 86% (41%) of Czechs would be uncomfortable having Roma (Vietnamese) neighbours.

Testing Attention Discrimination: A Lemon Dropping Market

Experiment consisted in sending emails expressing interest in arranging an apartment viewing.
Three fake applicants were generated:

- Vietnamese, Roma and white majority group

Each applicant had a name, email address and personal website

- Jiri Hajek (white maj), Phan Nguyen (Vietnamese), Gejza Horvath (Roma)
- Separate survey with different sample of landlords confirmed ethnicity-name associations.

Testing Attention Discrimination: A Lemon Dropping Market

TABLE S1 — CZECH RENTAL HOUSING MARKET – DESIGN OF THE EXPERIMENT

	No Information Treatment Email: name	**Monitored Information Treatment** Email: name and hyperlink to website Website: info about education, occupation, age, marital status, smoking		**Treatment with additional text in the email** Email: name, info about education, occupation, age, marital status, smoking	
		High school degree	College degree	High school degree	College degree
White majority name	X	X	X	X	X
Asian minority name	X	X	X	X	X
Roma minority name	X	X	X	X	X

Testing Attention Discrimination: A Lemon Dropping Market

Supplementary Figure 1 — Applicant's Personal Website Snapshot (Czech Rental Housing Market)

Testing Attention Discrimination: A Lemon Dropping Market

Table 1—Czech Rental Housing Market: Invitation Rates and Information Acquisition by Ethnicity, Comparison of Means

	White majority name (W) (1)	Pooled Asian and Roma minority name (E) (2)	Percentage point difference: W − E, (*p*-value) (3)	Asian minority name (A) (4)	Percentage point difference: W − A, (*p*-value) (5)	Roma minority name (R) (6)	Percentage point difference: W − R, (*p*-value) (7)	Percentage point difference: R − A, (*p*-value) (8)
Panel A. Invitation for a flat visit								
No Information Treatment ($n = 451$)	0.78	0.41	37 (0.00)	0.39	39 (0.00)	0.43	36 (0.00)	3 (0.57)
Monitored Information Treatment ($n = 762$)	0.72	0.49	23 (0.00)	0.49	23 (0.00)	0.49	23 (0.00)	0 (0.92)
Monitored Information Treatment[a] ($n = 293$)	0.84	0.66	18 (0.00)	0.71	13 (0.00)	0.62	21 (0.00)	−9 (0.20)
Monitored Information Treatment[b] ($n = 469$)	0.66	0.37	29 (0.00)	0.35	31 (0.00)	0.39	27 (0.00)	4 (0.51)
Treatment with additional text in the e-mail ($n = 587$)	0.78	0.52	26 (0.00)	0.49	29 (0.00)	0.55	23 (0.00)	5 (0.29)
Panel B. Information acquisition in the Monitored Information Treatment								
Opening applicant's personal website	0.33	0.41	−8 (0.03)	0.38	−5 (0.24)	0.44	−11 (0.01)	6 (0.15)
Number of pieces of information acquired	1.29	1.75	−0.46 (0.01)	1.61	−0.32 (0.09)	1.88	−0.59 (0.00)	0.27 (0.17)
At least one piece of information acquired	0.30	0.40	−10 (0.01)	0.37	−7 (0.12)	0.44	−13 (0.00)	7 (0.12)

Testing Attention Discrimination: A Lemon Dropping Market

Observation 1: Applicants with minority-sounding names are discriminated against.

- If no information about applicants is available, majority applicants are 90% more likely to be invited for an apartment viewing than minority applicants.

Testing Attention Discrimination: A Lemon Dropping Market

Observation 2: Landlords pay more attention to available information about applicants with a minority-sounding name relative to applicants with a majority-sounding name.

Observation 3: Landlords' invitation decision is responsive to the available information about minority applicants, while the same is not true about applicants with a majority-sounding name.

Testing Attention Discrimination: A Cherry Picking Market

The authors used the same names, but this time as applicants for jobs in response to job ad online.

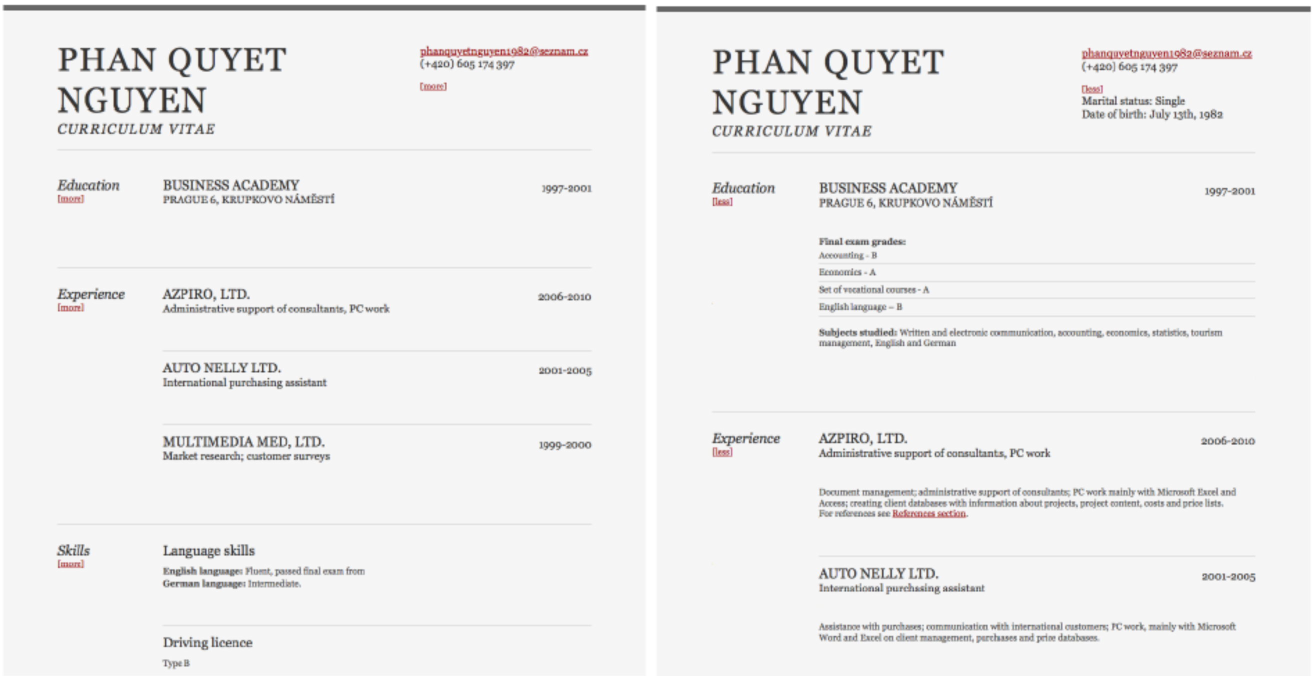
- Only implemented the Monitoring of Information Acquisition treatment

Email applying for a job included hyperlink to a resume in a personal website.

Testing Attention Discrimination: A Lemon Dropping Market

SUPPLEMENTARY FIGURE 2 — APPLICANT'S ONLINE RESUME, CZECH LABOR MARKET

Left Part: A Snapshot After Opening the Website (a Shorter Form), Right Part: A Snapshot After Expanding Education and Experience Categories



Testing Attention Discrimination: A Lemon Dropping Market

Table 4—Czech Labor Market: Invitation Rates and Information Acquisition by Ethnicity, Comparison of Means

	White majority name (W) (1)	Pooled Asian and Roma minority name (E) (2)	Percentage point difference: W − E, (p-value) (3)	Asian minority name (A) (4)	Percentage point difference: W − A, (p-value) (5)	Roma minority name (R) (6)	Percentage point difference: W − R, (p-value) (7)	Percentage point difference: R − A, (p-value) (8)
Panel A. Employer's response								
Callback	0.43	0.20	23 (0.00)	0.17	26 (0.00)	0.25	18 (0.01)	8 (0.22)
Invitation for a job interview	0.14	0.06	8 (0.03)	0.05	9 (0.03)	0.08	6 (0.18)	3 (0.46)
Invitation for a job interview[a]	0.19	0.09	10 (0.06)	0.09	10 (0.12)	0.10	9 (0.16)	1 (0.83)
Panel B. Information acquisition								
Opening applicant's resume	0.63	0.56	7 (0.22)	0.47	16 (0.03)	0.66	−3 (0.69)	19 (0.01)
Acquiring more information about qualification[a]	0.16	0.10	6 (0.27)	0.06	10 (0.12)	0.14	2 (0.73)	8 (0.24)
Acquiring more information about other characteristics[a]	0.18	0.18	0 (0.92)	0.19	−1 (0.85)	0.18	0 (0.99)	1 (0.85)

Testing Attention Discrimination: A Lemon Dropping Market

Observation 4: Applicants with minority-sounding names are discriminated against in the labor market.

- Majority applicants are 180% (75%) more likely to be invited for a job interview than Asian (Roma) applicants.

Observation 5:

- Employers are 34 percent more likely to read a resume provided by majority applicants than Asian applicants.
- Conditional on opening a resume, employers more closely inspect qualifications of majority than Asian applicants.
- Small differences in the likelihood of opening a resume/depth of inspection between majority and Roma applicants.

Testing Attention Discrimination: Lessons to learn

When a information is revealed matters!

- The later key attributes are revealed, the more attention/relevance education or qualifications are in the DM's process
- e.g. 'blind' auditions increased the likelihood of female musicians being selected by 30% in US symphony orchestras (Goldin and Rouse, 2000).
- Name blind CVs? (Aslund and Skans 2012)

May want to provide early signals early, rather than hoping the selector will pick that up.

Personal Contact/Audit Field Experiments

Personal Contact/Audit studies differ from Written Contact studies by the fact that face-to-face contact exists between the employer and the candidate

Often candidates are trained actors, who will likely be able to give a more convincing/consistent performance than non-actors.

- Training consists in giving the full set of actors a similar background
- They should behave as closely as possible so as to make [race/gender/other] the only differing characteristic

Personal Contact/Audit Field Experiments

Advantages:

- Cleaner effect of characteristics that are not 100% verifiable in a CV (e.g. ethnicity)
- Richer dataset: job offer, salary.

Disadvantages:

- Unclear the extent to which actors are truly “equivalent”
- Not double-blind
- Potential experimenter demand effect: actors know the purpose of the study
- Very costly, so limited scope to vary characteristics

Castillo, Petrie, Torero and Vesterlund (2013)

CPTV conduct an audit field experiment to test for gender discrimination in taxi fares in in Lima, Peru.

- The taxi market in Lima is highly competitive (200,000 taxis in a city with 7.7 million inhabitants.
- By comparison NYC has 53,000 licensed taxis in a population of 8.3 million
- Taxi driver earnings are 30-50 soles for a 13-hour day $\approx$ minimum wage
- Vast majority of Lima residents use a taxi (only 17% own their car)

Castillo, Petrie, Torero and Vesterlund (2013)

This is a market in which both male and female passengers are highly experienced.

- Taxi fares account for roughly 8.8% of Lima households' budgets

The object of negotiation is well defined: a fare to get from current location to location X.

- Lima taxis do not have meters and there are no formally-defined zones
- The fare must be agreed by face-to-face negotiation.
- No tipping, so fare accounts for the entire price.

The challenge is to design a gender-neutral negotiation strategy.

Castillo, Petrie, Torero and Vesterlund (2013)

CPTV trained six men and women to be ‘taxi passengers’.

- Passengers are instructed to negotiate for and travel along a number of routes.
- A route is: A $\rightarrow$ B, B $\rightarrow$ C and C $\rightarrow$ A
- They also considered reverse directions as a control.

Castillo, Petrie, Torero and Vesterlund (2013)

At each location, "passengers" hailed a taxi, approached the passenger window and asked: "How much would it cost to go to X?"

After the taxi driver quotes a price, the passenger follows a fixed-offer bargaining script:

- For any price quoted by the taxi driver, the "passenger" responds with p_{max}.
- Always respond with p_{max} until either the taxi driver accepts, or drives away

Castillo, Petrie, Torero and Vesterlund (2013)

Note that in this bargaining script, the only person who changes price is the taxi driver.

- Therefore the driver is the only person who chooses to terminate the negotiation.

If the first price quoted by the taxi driver is $\leq p_{max}$, or if the driver subsequently accepts p_{max} the negotiation ends.

If the driver refuses, the "passenger" would step away from the street, take out a mobile phone and pretend to be receiving a call.

- To clear the street from other taxi drivers who would be waiting for negotiation to break down.

Castillo, Petrie, Torero and Vesterlund (2013)

"Passengers" who could not agree a trip for p_{max} took a taxi to the next location at a possibly higher price and start again.

- Those observations are excluded
- CPTV did this to ensure all "passengers" would get data in all locations.

p_{max} was chosen to be low enough to provoke counter-offers, but high enough to be acceptable.

Castillo, Petrie, Torero and Vesterlund (2013)

This design has several useful features:

- The bargaining strategy is widely used (it is familiar to drivers)
- It allows CPTV to collect several key data:
  - Initial offer
  - Number of counter-offers
- Easy to be consistent across "passengers"

The experiment was conducted in central, business locations between 8am and 1pm, Monday through Friday.

- The objective of travel should be the same for men and women
- Drivers have similar outside options

Castillo, Petrie, Torero and Vesterlund (2013)

Table 1
Distribution of initial prices by maximum-acceptable offer.[a]

Initial price	Maximum-acceptable offer				Total
	3	4	5	6	
3	7	0	0	0	7
4	18	9	0	0	27
5	**51**	100	9	0	160
6	20	**212**	31	5	268
7	19	128	51	66	264
8	3	57	**107**	**98**	265
9	1	5	21	20	47
10	0	4	24	20	48
12	0	0	1	1	2
13	0	0	0	1	1
15	0	0	0	1	1
Total	119	515	244	212	1090
Average initial price (SD)	5.3 (1.2)	6.3 (1.0)	7.7 (1.2)	8.0 (1.1)	6.8 (1.4)
Rejection rate	55.5%	63.9%	73.4%	50.5%	62.5%

Note: The highlighted bold entries indicate the modal price for each maximum-acceptable offer.

[a] Included in the 6 soles maximum-acceptable offer routes are also two observations where passengers (both male) incorrectly used a maximum-acceptable price of 7 soles. None of our results are influenced by this inclusion.

Castillo, Petrie, Torero and Vesterlund (2013)

Table 2

Distribution of negotiation outcomes (row percentage in parentheses).

	Acceptances	Rejections	Renegotiations	Total
Round 1	221 (20)[a]	303 (28)	566 (52)	1090
Round 2	136 (24)	271 (48)	159 (28)	566
Round 3	44 (28)	90 (57)	25 (16)	159
Round 4	7 (28)	16 (64)	2 (8)	25

[a] As seen in Table 1, 30 of these round-1 agreements result from the driver proposing an initial price equaling the maximum-acceptable offer.

Castillo, Petrie, Torero and Vesterlund (2013)

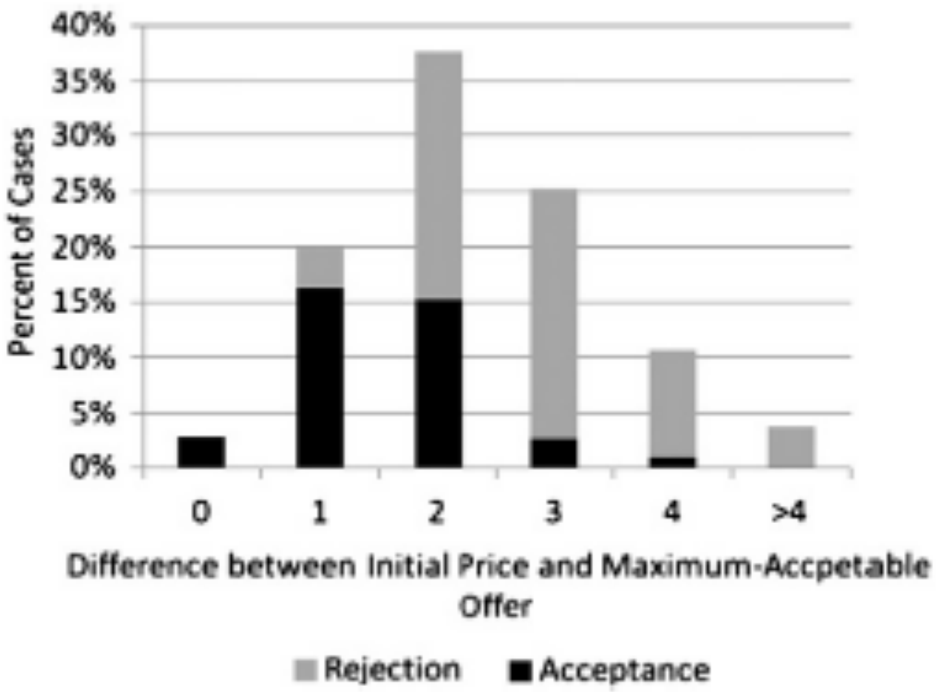


Fig. 3. Bargaining outcome conditional on difference between, initial price and maximum-acceptable offer.

Castillo, Petrie, Torero and Vesterlund (2013)

Table 3
OLS regressions on initial and last acceptable price.

Variables	(1)	(2)
	Initial price	Final acceptable price
Male	0.22	0.32
	(0.02)	(0.01)
Constant	6.22	5.31
	(0.00)	(0.00)
Observations	1090	1090

p-Values in parentheses. Date, time, route and passenger random effects.

Castillo, Petrie, Torero and Vesterlund (2013)

Table 4
OLS regressions on prices across rounds by gender.[a]

Variables	(1) Initial price	(2) Second price	(3) Third price
Male	0.22	0.24	0.06
	(0.02)	(0.08)	(0.73)
Constant	6.22	5.46	5.56
	(0.00)	(0.00)	(0.00)
Observations	1090	566	159

p-Values in parentheses. Date, time, route and passenger random effects.

[a] Because men are rejected more frequently than women, the women in our study are spending more time riding taxis which in turn implies that a larger fraction of our negotiations are done by men. Using routes as the unit of observation, the average proportion of observations produced by men is 55.7% (SD 17.4, median 55).

Castillo, Petrie, Torero and Vesterlund (2013)

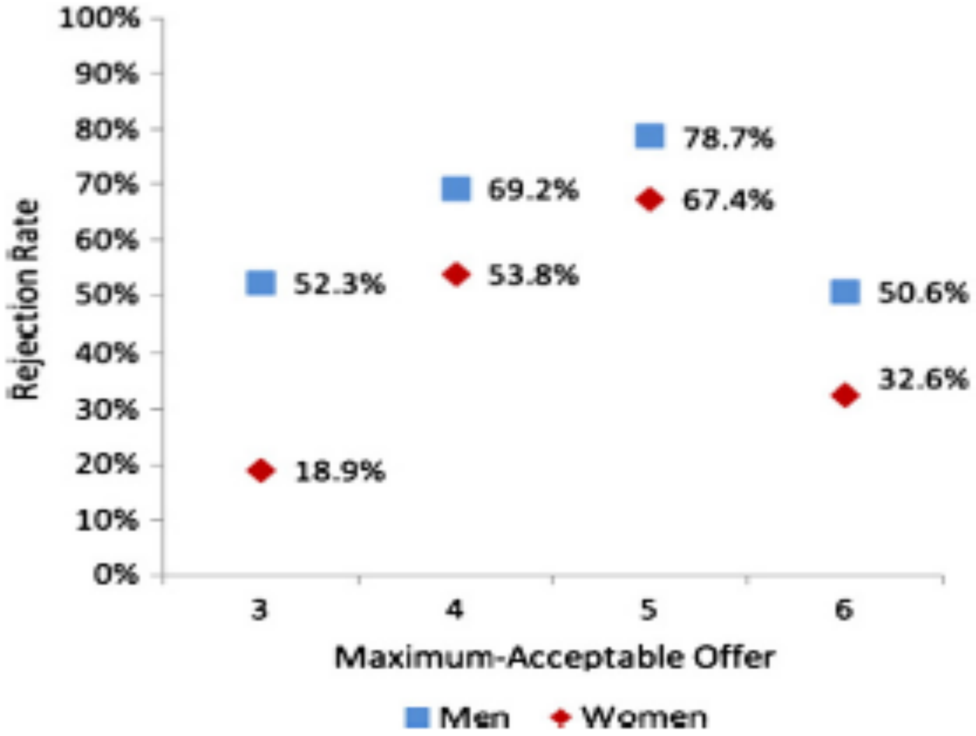


Fig. 4. Rejection rate by maximum-acceptable offer (using matching methods).

Castillo, Petrie, Torero and Vesterlund (2013)

How to distinguish between taste-based discrimination and statistical discrimination?

- Exploit different market conditions during the day.
- Peak and off-peak times imply different "costs of discrimination".

The cost of taste-based discrimination is:

- Greater late in the morning when there are fewer customers.
- Smaller early in the morning when there are more high-valuation female passengers.

Castillo, Petrie, Torero and Vesterlund (2013)

In contrast, the cost of statistical discrimination is:

- Higher early in the morning when passengers are more homogeneous.
- Lower late in the morning when female passengers potentially have a higher p_{max}.

Castillo, Petrie, Torero and Vesterlund (2013)

Table 8

OLS regressions on initial price over course of the morning (study 1).

Variables	(1)	(2)	(3)	(4)	(5)
	8 am	9 am	10 am	11 am	12 pm
Male	0.11	0.02	0.35	0.21	0.24
	(0.49)	(0.86)	(0.05)	(0.24)	(0.32)
Constant	7.03	6.46	6.06	6.01	6.40
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	221	237	246	268	118

p-Values in parentheses. The period 8–8:59 am is recorded as 8 am.
Date, route and passenger random effects.

Castillo, Petrie, Torero and Vesterlund (2013)

In short, CPTV find strong evidence for statistical discrimination against males, rather than taste-based discrimination.

Discrimination is driven by the perception that males have higher willingness to pay for a taxi ride than females

That means the opportunity cost of discriminating is highest in off-peak times, when travel motives are not work-related.