Rationally Inattentive Statistical Discrimination: Arrow Meets Phelps

Federico Echenique Anqi Li
UC Berkeley Virginia Tech

Encuentro SEU, December 22nd 2022

Statistical Discrimination

Members of a demographic group are discriminated against due to "rational" beliefs about the group.

Contrast with prejudice or animosity.

Two canonical models:

- Arrovian model of coordination failure (Arrow, 1971; Coate and Loury, 1993).
- Phelpsian model of information heterogeneity (Phelps, 1972; Aigner and Cain, 1977).

This Paper: Arrow meets Phelps

Endogenize information in an arrovian setting of costly unobservable investments: Information (or attention) is costly (Sims, 2003; Matějka–McKay, 2015).

This Paper: Arrow meets Phelps

An employer may save on costly information by monitoring Group 1 more favorably than Group 2.

Meaning succesfull investments by G1 are more easily detected (and thus rewarded) than G2.

Imagine (as an illustration) observing first a signal about G1, and only if that signal is negative, get a signal about G2.

When G1's are more likely to invest, this can be a good idea.

This Paper: Arrow meets Phelps

Group 1 is therefore incentivized to undertake costly investment.

Group 2 is discouraged as they need to pass a higher informational bar ("stronger signal") before their investment is recognized.

Information about G2 is less valuable than about G1, as they conform to stereotype — closing a vicious circle.

This Paper: Arrow meets Phelps

We'll have multiple equilibria, as in Arrow.

But an eqm. selection story: situations where the most profitable eqm. to the employer is discriminatory.

+ Informational heterogeneity as in Phelps, but endogenously generated.

Hence,

This Paper: Extensions

Affirmative action — quotas vs. subsidies.

Multi-task version of our model generates occupational segregation and stereotypes.

Social Psychology Research on Implicit Bias and Stereotype

Folk wisdom: Deficits in attention capacity trigger implicit biases and stereotypes (Greenwald et al., 1998)

Rich evidence in numerous contexts:

- Labor: Chugh (2004), Bertrand et al. (2005)
- Criminal justice: Eberhardt (2020)
- Education: Warikoo et al. (2017)
- Healthcare: Chapman et al. (2013)

Lack a theory that formalizes the causal link.

The Model

- A principal and two agents: Michael (m) and Wendy (w)
- Agent $i \in \{m, w\}$ exerts effort $\mu_i \in \{\underline{\mu}, \overline{\mu}\}$ at a cost $C(\mu_i)$;
- $0 < \underline{\mu} < \overline{\mu} < 1$ and $C(\underline{\mu}) = 0 < C(\overline{\mu}) = C$
- Effort choice is unobserved to the principal.
- Effort generates a random productivity $\tilde{\theta}_i$ equal to 1 with prob. μ_i and 0 with prob. $1 - \mu_i$

The Model

- Principal chooses either Michael or Wendy to promote.
- m and w get payoff $= 1$ from being promoted.
- Principal gets more from promoting productive agents (payoff $= \theta_i$).
- Principal acquires costly information about θ_is choosing signal structure $\pi : \{0,1\}^2 \to \Delta(S)$
- Promotion: $a : S \to \Delta(\{0,1\})$, where $a = 1$ (resp. $a = 0$) means that m (resp. w) is promoted

The Principal's Problem (Cont'd)

For any given profile $\boldsymbol{\mu} = (\mu_m, \mu_w)$ of efforts, the principal's expected payoff is

$$\underbrace{\mathbb{E}\left[\tilde{a}\Delta\tilde{\theta} \mid \boldsymbol{\mu}, \pi, a(\cdot)\right] + \mu_w}_{\text{expected revenue}} - \lambda \cdot I(\pi \mid \boldsymbol{\mu})$$

- $\Delta\theta = \theta_m - \theta_w$: differential productivity between m and w
- $\lambda > 0$: attention cost parameter
- $I(\pi \mid \boldsymbol{\mu})$: mutual information cost

Game

A simultaneous game where

- The principal specifies $(\pi, a(\cdot))$
- Agents choose μ_is

PSBNE as the solution concept

Simplify the Principal's Problem

W.l.o.g. consider signals of form $\pi : \{-1, 0, 1\} \to \Delta(\{0, 1\})$:

- Recommend an agent for promotion based on $\Delta\theta$
- Recommendation is strictly obeyed by the principal

Key Concepts

*π is **impartial** if the prob. of promoting an agent depends only on differential productivity, not on agents' identities.*
*Otherwise π is **discriminatory**.*

*An equilibrium is **impartial** if the equilibrium signal is impartial.*
*Otherwise it is **discriminatory**.*

Summary of Results

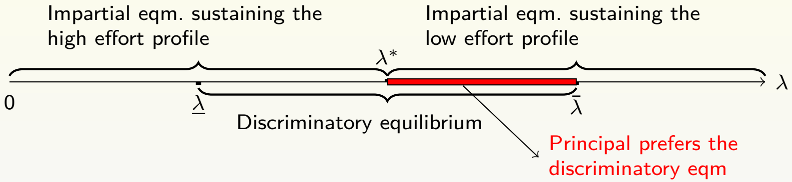

Figure 1: Equilibrium regimes

[redacted]

The following are true if $\overline{\mu} + \underline{\mu} > 1$ *and* $\frac{C}{\overline{\mu} - \underline{\mu}} < \frac{\overline{\mu}(1-\overline{\mu})}{\overline{\mu} + \underline{\mu} - 2\overline{\mu}\underline{\mu}}$.

(i) *An impartial equilibrium always exists, and it is unique if and only if* $\lambda \neq \lambda^*$ *for some* $\lambda^* > 0$*. When unique, the impartial equilibrium sustains* $(\overline{\mu}, \overline{\mu})$ *if* $\lambda < \lambda^*$ *and* $(\underline{\mu}, \underline{\mu})$ *if* $\lambda > \lambda^*$.

(ii) *A discriminatory equilibrium sustaining* $(\overline{\mu}, \underline{\mu})$ *uniquely exists if and only if* $\lambda \in [\underline{\lambda}, \overline{\lambda}]$ *for some* $0 < \underline{\lambda} < \overline{\lambda} < +\infty$.

(iii) $\underline{\lambda} < \lambda^*$ *always holds.* $\lambda^* < \overline{\lambda}$ *holds under further regularity conditions.*

Statement of Results (Cont'd)

Let everything be as in Theorem 1, and suppose that $\lambda^* < \overline{\lambda}$*. Then the most profitable equilibrium is discriminatory iff* $\lambda \in (\lambda^*, \overline{\lambda}]$*.*

Intuition for Theorem 1

The optimal signal for $(\overline{\mu}, \underline{\mu})$ favors *m* over *w* unless *w* is strictly more productive than *m*

Table 1: Optimal signal structure for $\boldsymbol{\mu} = (\overline{\mu}, \underline{\mu}) = (.8, .6)$, $\lambda = .3$

$\Delta\theta$	1	0	-1
$\mathbb{P}(\Delta\theta \mid \boldsymbol{\mu})$	.32	.56	.12
$\pi(\Delta\theta)$	.98	.74	.09

In this way, the principal can afford to be rationally inattentive but still does a decent job in selecting the most productive agent

Intuition for Theorem 1 (Cont'd)

By exerting high effort rather than low effort, w can increase her promotion prob. by

$$(\overline{\mu} - \underline{\mu})[\overline{\mu}(\pi(1) - \pi(0)) + (1 - \overline{\mu})(\pi(0) - \pi(-1))] = .081$$

The number for m is

$$(\overline{\mu} - \underline{\mu})[(1 - \underline{\mu})(\pi(1) - \pi(0)) + \underline{\mu}(\pi(0) - \pi(-1))] = .098$$

If $C \in [.081, .098]$, then m exerts high effort and w low effort

Intuition for Theorem 2

When attention is scarce,

- The only way to maintain impartiality is to use a noisy signal that provides uniformly low incentives to agents
- Better use a discriminatory signal to save on the attention cost and to boost the revenue

When attention is abundant,

- The discriminatory equilibrium has a cost advantage and a revenue disadvantage compared to the impartial equilibrium
- The revenue concern is always dominant

Implication I: Varying Effectiveness of De-biasing Programs

Our model predicts a nonmonotonic relation between the attention cost parameter and the equilibrium degree of discrimination

Speak to the varying effectiveness of de-biasing programs that modulate the attention channel:

- Changes in the foot pursuit policy of the Oakland Police Department (Eberhardt, 2020)
- Nextdoor Neighbor's addition of frictions to communal crime reporting
- Implicit bias training programs (Greenwald and Lai, 2020)

Implication II: Gender and Racial Gap in Subjective Performance Evaluation

Our model predicts that minorities are rated more harshly, unless they truly are strictly more productive than majorities

Speak to the gender and racial gap in subjective performance evaluation:

- Women are held to higher performance standards, and face increased scrutiny and shifting criteria (Wynn and Correll, 2018)
- They tend to get shorter, more vague, and less constructive feedback (Correll and Simard, 2016; Mackenzie et al., 2019; Jampol and Zayas, 2021)

Implication III: Welfare

Equilibria cannot be Pareto ranked:

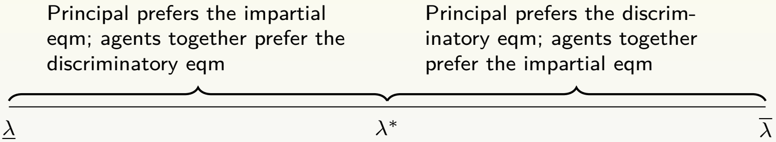

Figure 2: Welfare regimes

Formal Analysis

For any π, write $\overline{\pi}$ for the average prob. that m is recommended for promotion, X for $\pi(1) - \pi(0)$, and Y for $\pi(0) - \pi(-1)$

Impartiality $\Longleftrightarrow \pi(0) = 1/2$ and $X = Y$

Write c for $C/\Delta\mu$, A for $\overline{\mu}(1 - \underline{\mu})$, and B for $\underline{\mu}(1 - \overline{\mu})$

[redacted]

$\overline{\mu} + \underline{\mu} > 1$ *and* $c < \overline{\mu}(1 - \overline{\mu})/(A + B)$.

Optimal Signal Structure

(i) *The optimal signal for* $(\overline{\mu}, \overline{\mu})$ *or* $(\underline{\mu}, \underline{\mu})$ *is nondegenerate and impartial, satisfying*

$$\overline{\pi} = \pi(0) = 1/2 \text{ and } X = Y = g(\lambda),$$

where $g > 0$ *and is decreasing in* λ.

(ii) *The optimal signal for* $(\overline{\mu}, \underline{\mu})$ *is degenerate if and only if* $\lambda \geq \check{\lambda} = (\ln(A/B))^{-1} > 0$. *When nondegenerate, the signal is discriminatory, satisfying*

$$\overline{\pi} = \pi(0) \in (1/2, 1) \text{ and } X = f(\lambda) < Y = \frac{A}{B} f(\lambda),$$

where $f > 0$ *and is decreasing in* λ.

Optimal Effort Choice

Fix any signal structure π. For any $\mu_w \in \{\overline{\mu}, \underline{\mu}\}$, m prefers to exert high effort rather than to exert low effort iff

$$(1 - \mu_w)X + \mu_w Y \geq c. \tag{IC$_m$}$$

For any $\mu_m \in \{\overline{\mu}, \underline{\mu}\}$, w prefers to exert high effort rather than to exert low effort iff

$$\mu_m X + (1 - \mu_m) Y \geq c. \tag{IC$_w$}$$

Equilibrium

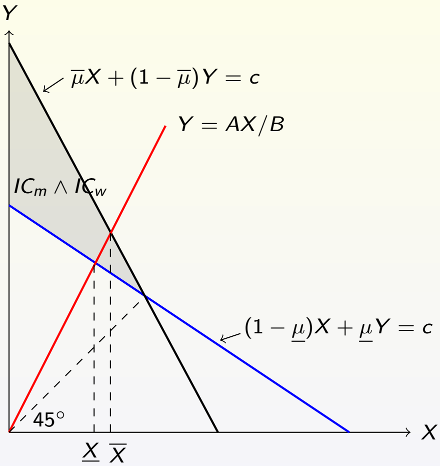

Figure 3: The discriminatory case

Discussions

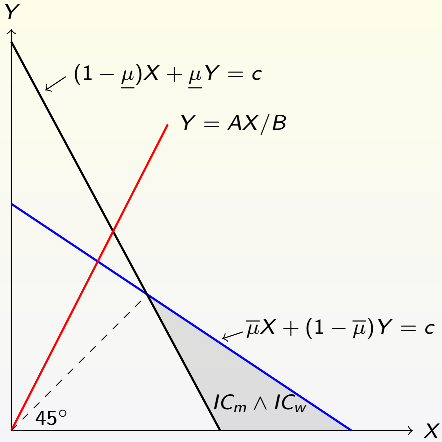

Figure 4: Discriminatory equilibrium doesn't exist if $\overline{\mu} + \underline{\mu} < 1$

Discussions (Cont'd)

$c < \frac{\overline{\mu}(1-\overline{\mu})}{A+B} \Longrightarrow$ the game has a unique equilibrium that sustains the high effort profile when $\lambda \approx 0$

RI turns agents' competition for the promotion opportunity into a competition for the principal's attention

Agenda

- Model
- Extensions
- Literature

Quota

Suppose the principal must promote the agents with equal probability on average:

$$\overline{\pi} = 1/2 \tag{Q}$$

When $\overline{\mu} + \underline{\mu} > 1$, a strategy profile constitutes an equilibrium of the game with (Q) if and only if it is an impartial equilibrium of the baseline game.

Implication

A vast literature studies the channels through which affirmative action quotas operate, as well as the duration of their effects (Holzer and Neumark, 2000; Fang and Moro, 2011; Doleac, 2021; Dianat et al., 2022)

In the current context,

- The quota operates only through eliminating the discriminatory equilibrium
- In the case where the discriminatory equilibrium is the most profitable to the principal, the effect of the quota is likely to be reversed after its removal

Optimal Signal Under Quota Constraint

The optimal signal structure for $(\overline{\mu}, \underline{\mu})$ *subject to (Q) uniquely exists and satisfies* $\pi(0) < 1/2$ *and* $X > Y > 0$.

Reverse the situations faced by m and w in order to satisfy the quota constraint

Equilibrium Under Quota Constraint

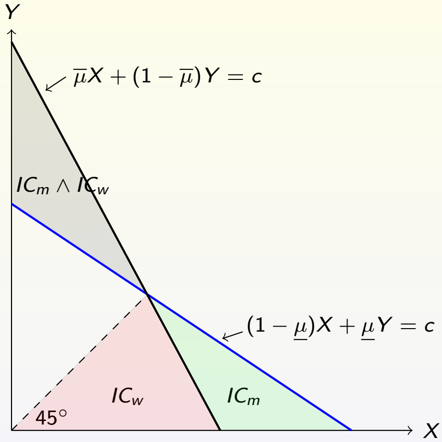

Figure 5: $(\overline{\mu}, \underline{\mu})$ cannot be sustained in any equilibrium under (Q)

Subsidy

Suppose that the principal receives a subsidy $s \geq 0$ for hiring w

An equilibrium is equitable if agents exert the same level of effort and are promoted with equal probability on average.

Under the assumption that $\overline{\mu} + \underline{\mu} > 1$, *there exists no* $s > 0$ *such that the corresponding game sustains* $(\overline{\mu}, \overline{\mu})$ *in an equilibrium. If, in addition,* $\underline{\mu} > 1/2$, *then then there exists no* $s > 0$ *such that the corresponding game has an equitable equilibrium.*

Quota vs. Subsidy

A sizable economic literature dating back to Weitzman (1974) examines the differences between price and quantity regulations

In the current context, quota and subsidy change the principal's problem to

$$\max_{\pi, a(\cdot)} \mathbb{E}\left[\tilde{a}(\Delta\tilde{\theta} - s) \mid \boldsymbol{\mu}, \pi, a\right] - \lambda \cdot I(\pi \mid \boldsymbol{\mu})$$

- In the case of quota,

$$s \begin{cases} = 0 & \text{if (Q) is automatically satisfied in the baseline model,} \\ > 0 & \text{otherwise.} \end{cases}$$

- In the case of subsidy, $s \geq 0$ is determined exogenously and rigidly by the authority

Optimal Signal Under Subsidy

[redacted]

Fix any effort profile $\boldsymbol{\mu}$ *and any positive level* $s > 0$ *of subsidy. If the optimal signal structure for* $\boldsymbol{\mu}$*, given the subsidy, satisfies (Q), then it must also satisfy* $\pi(0) < 1/2$ *and* $X > Y > 0$*.*

Equilibrium Under Subsidy

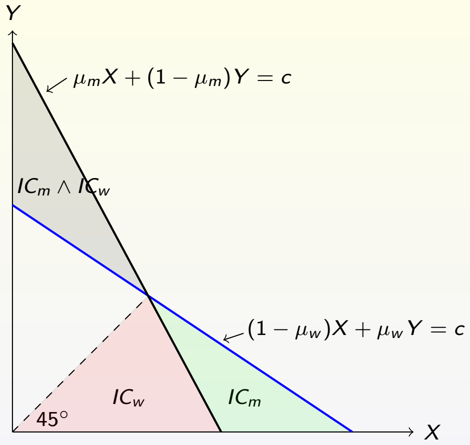

Figure 6: When $\mu_m + \mu_w > 1/2$, a signal structure with $X > Y$ cannot be incentive compatible for both agents

Multiple Tasks and Skills

There are two tasks that need to be performed: $t = 1, 2$

Each task t arrives with prob. $\alpha^t \in (0, 1/2]$. The two tasks never arrive simultaneously

Each agent i can undertake costly investments in task-specific skills: $\mu_i^t \in \{\underline{\mu}, \overline{\mu}\} \; \forall t$

μ_i^t generates θ_i^t equal to 1 with prob. μ_i^t and 0 with prob. $1 - \mu_i^t$, at a cost $C^t(\mu_i^t)$ where $C^t(\underline{\mu}) = 0 < C^t(\overline{\mu}) = C^t$

If agent i is chosen to perform task t, then he earns a prize β^t and delivers a benefit θ_i^t to the principal

Game

The signal that screens agents for task t is $\pi^t : \{-1, 0, 1\} \to [0, 1]$

A simultaneous game where

- The principal specifies π^t, $t = 1, 2$
- Each agent i specifies $\boldsymbol{\mu}_i = (\mu_i^1, \mu_i^2)$

After that, the task that needs to be performed arrives, and agents are screened by the pre-specified signal

Key Concept

*Call an equilibrium **non-specialized** if both agents adopt the same investment strategy. Call an equilibrium **specialized** if one agent invests in skill 1 and the other agent invests in skill 2.*

Interpretation:

- Label one task as "traditionally male" and the other as "traditional female." Screen m and w favorably for their respective tasks
- Agents invest in the skills they are screened favorably for, which reinforces the use of specialized screening

Regularity Condition

Define $c^t = C^t/(\alpha^t \beta^t \Delta\mu)$. Assume w.l.o.g. that $c^1 \leq c^2$

Let $\lambda^*(c)$, $\overline{\lambda}(c)$ and $\underline{\lambda}(c)$ denote the cutpoints in λ as decreasing functions of c

Suppose the regularity conditions stated in Theorem 1 hold for $t = 1, 2$, hence $0 < \underline{\lambda}(c^t) < \lambda^*(c^t) < \overline{\lambda}(c^t)$ for $t = 1, 2$.

The following statements are true under Assumption 2.

(i) *A non-specialized equilibrium always exists and is generically unique. When unique, the equilibrium induces both agents to invest in both skills when* $\lambda < \lambda^*(c^2)$, *no agent to invest in any skill when* $\lambda > \lambda^*(c^1)$, *and both agents to invest in skill 1 but not skill 2 when* $\lambda \in (\lambda^*(c^2), \lambda^*(c^1))$.

(ii) *A specialized equilibrium exists if and only if*

$$\frac{c^1}{c^2} \geq \frac{\overline{\mu}(1-\overline{\mu})}{\underline{\mu}(1-\underline{\mu})} \text{ and } \lambda \in [\underline{\lambda}(c^1), \overline{\lambda}(c^2)].$$

Whenever a specialized equilibrium exists, it is unique.

Results (Cont'd)

Let everything be as in Theorem 5, and suppose that $\alpha^1 = \alpha^2$.

(i) *When the game has a specialized equilibrium and a non-specialized equilibrium in which both agents invest in both skills, the non-specialized equilibrium is the most profitable.*

(ii) *When the game has a specialized equilibrium and a non-specialized equilibrium in which no agent invests in any skill, the specialized equilibrium is the most profitable.*

(iii) *When the game has a specialized equilibrium and a non-specialized equilibrium in which both agents invest in skill 1 but not skill 2, the specialized equilibrium is the most profitable.*

Implications

A RI theory of occupational segregation and stereotypes

Speak to stylized facts, including:

- Occupational segregation exists even within narrowly defined firms or industries (Blau and Kahn, 2017)
- Women are evaluated based on personalities and likeabilities, and are under-rewarded for traits associated with men such as taking charge or being visionary (Bohnet et al., 2016; Correll et al., 2020)
- Stereotypical performance evaluation is cited as a culprit for women's underrepresentation in STEM fields (Moss-Racusin et al., 2012; Lavy and Sand, 2018)

Agenda

- Model
- Extensions
- Literature

Literature

Theory of statistical discrimination:

- Arrow (1971, 1998), Coate and Loury (1993), Fang (2001), Chaudhuri and Sethi (2008)
- Phelps (1972), Aigner and Cain (1977), Chambers and Echenique (2021), Escudé et al. (2022)
- Borjas and Goldberg (1978), Lundberg and Startz (1983)
- Bordalo et al. (2016), Bohren et al. (2019), Heidues et al. (2019)
- Fang and Moro (2011), Onuchic (2022)

Literature (Cont'd)

Rational inattention:

- Yang (2015), Dessein et al. (2016), Hu et al. (2019), Ravid (2020), Yang (2020)
- Ambuehl (2017), Dean and Neligh (2017), Novák et al. (2021)
- Matějka and McKay (2015), Matveenko and Mikhalishchev (2021)

Attentional discrimination:

- Bartoš et al. (2016), Che et al. (2020), Fosgerau et al. (2021)
- Glover et al. (2017), Huang et al. (2022)

Social categorization:

- Allport (1954), Devine (1998); Gilbert and Hixon (1991), Greenwald and Banaji (1995), Macrae and Bodenhausen (2000)
- Fryer and Jackson (2008)

Literature (Cont'd)

Incentive theory:

- Alchian and Demsetz (1972), Li and Yang (2020)
- MacLeod (2003), Taylor and Yildirim (2011); Prendergast and Topel (1993, 1996)

Biased contest and affirmative action:

- Lazear and Rosen (1981), O'Keeff et al (1984); Drugov and Ryvkin (2017), Fu and Wu (2020)
- Dechenaux et al. (2015), Chowdhury et al. (2020)

Concluding Remarks

RI statistical discrimination where Arrow meets Phelps:

- Biased attention allocation, beliefs, and differing human capital investments among ex-ante identical agents, in the most profitable equilibrium to the principal
- A causal link between limited attention and implicit bias as postulated by the social psychology literature
- New insights into various labor market outcomes, as well as policy designs to curtail discrimination

Future directions:

- Dynamics
- Systemic attentional discrimination

Thank You