The Dynamics of Discrimination: Theory and Evidence

J. Aislinn Bohren, Alex Imas, Michael Rosenberg
AER, 2019

Presentation: Ehsan Sabouri
August 5, 2019

Introduction

- Due to endogeneity issues, many researchers have employed field experiments to study discrimination. Field experiments have been successful in causally identifying the incidence of discrimination, but most cannot identify the source of this discrimination (**Bertrand and Duflo 2016**).
- One notable exception is **List (2004)**, who documents that minorities receive inferior initial and final offers when bargaining in a market for sports cards.
- This paper findings shed light on the mechanism behind previously documented discrimination reversals. for example,**Petersen and Saporta (2004)** find discrimination against women at the initial hiring stage for promotable jobs, but conditional on being hired, they find that women are more likely to be promoted.

Description of Forum

Mathematics Stack Exchange

- field experiment on an online Q&A mathematics forum, Mathematics Stack Exchange. Mathematics Stack Exchange is part of a family of forums where, in 2017 alone, 3,517,799 questions were asked and 4,299,077 answers were provided.(over 10 million registered users)

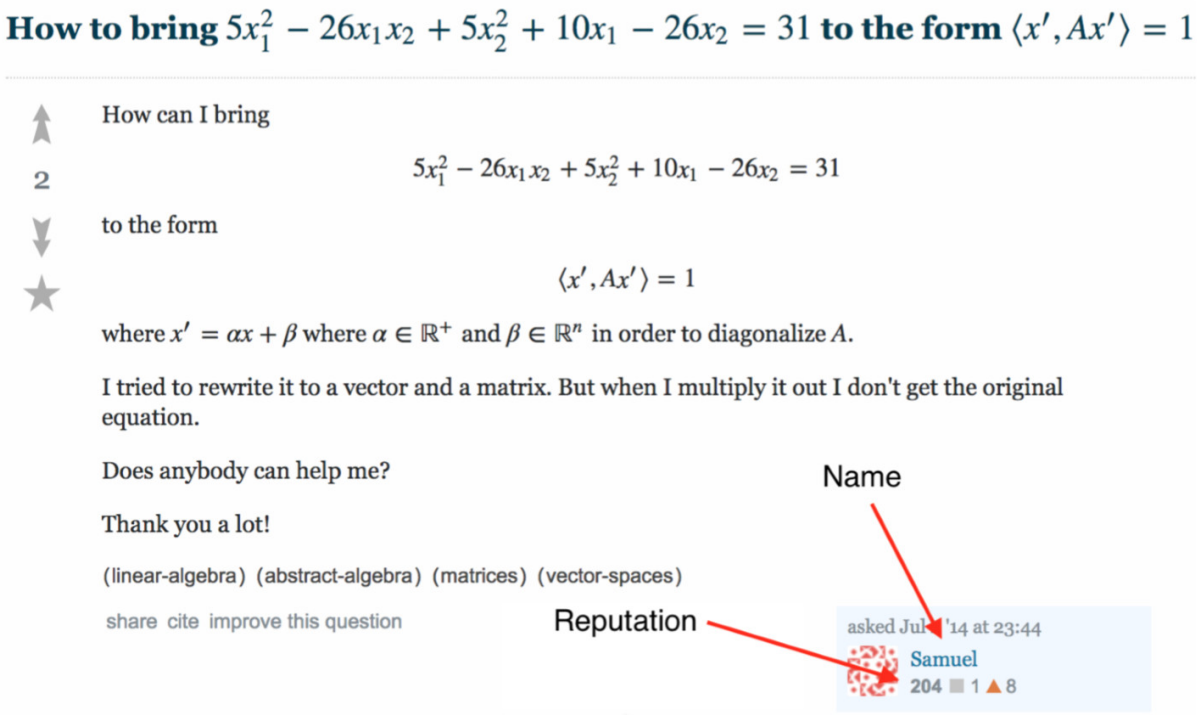


Figure 1: Question Post

Description of Forum(continued)

Mathematics Stack Exchange

- An upvote earns five reputation points on questions and ten reputation points on answers, while a downvote deducts two reputation points for both questions and answers.
- Reputation unlocks privileges, such as the ability to edit and comment on others posts or tag questions as duplicates. It can also be used as a currency through the assignment of "bounties".

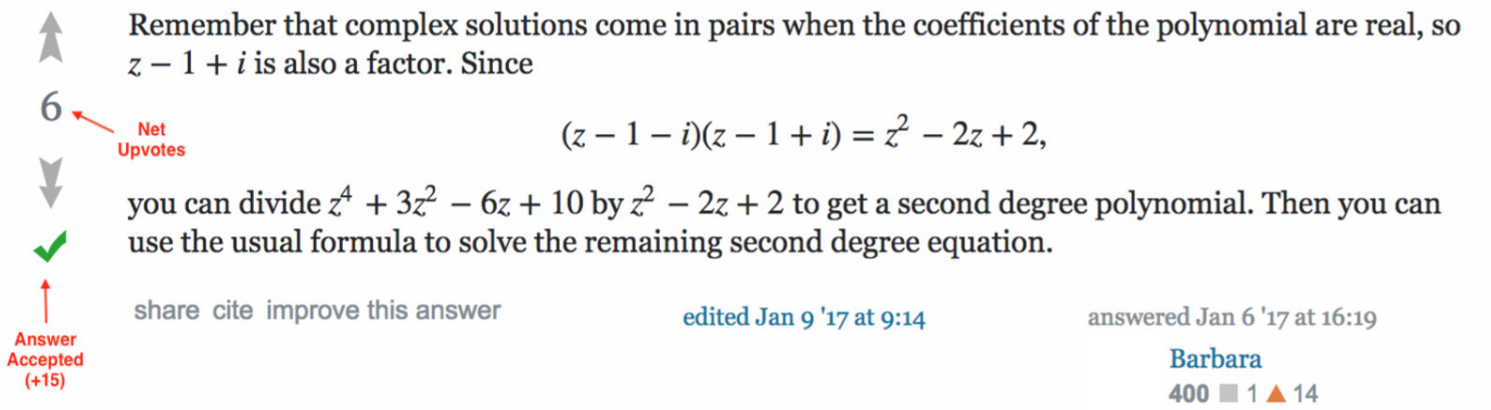


Figure 2: Answer Post

A Dynamic Model of Discrimination

Model Setting

Worker

Wrokers attribute :

- Each worker who has observable group identity $g \in \{F, M\}$
- unobservable ability $a \sim N(\mu_g, 1/\tau_a)$, with mean $\mu_g \in \mathbb{R}$ and precision $\tau_a > 0$.

The worker completes a sequence of tasks $t = 1, 2, \ldots$
Each task has hidden quality $q_t = a + \varepsilon_t$ where $\varepsilon_t \sim N(0, 1/\tau_\varepsilon)$ is an independent random shock with precision $\tau_\varepsilon > 1$.

Ability is fixed across time, and higher ability generates higher expected quality.

Model Setting

Evaluators

- A set of evaluators assess the worker's performance. For simplicity, assume that there is one evaluator per task, who reports evaluation $v_t \in \mathbb{R}$.
- Before evaluating task t, the evaluator observes:
  - the worker's gender g
  - evaluations on past tasks $h_t = (v_1, \ldots, v_{t-1})$, where $h_1 = \emptyset$
  - signal $s_t = q_t + \eta_t$ of quality of the current task, where $\eta_t \sim N(0, 1/\tau_\eta)$ is an independent random shock with precision $\tau_\eta > 0$

Model Setting

Evaluators

- A set of evaluators assess the worker's performance. For simplicity, assume that there is one evaluator per task, who reports evaluation $v_t \in \mathbb{R}$.
- Before evaluating task t, the evaluator observes:
  - the worker's gender g
  - evaluations on past tasks $h_t = (v_1, \ldots, v_{t-1})$, where $h_1 = \emptyset$
  - signal $s_t = q_t + \eta_t$ of quality of the current task, where $\eta_t \sim N(0, 1/\tau_\eta)$ is an independent random shock with precision $\tau_\eta > 0$
- An evaluator's type θ_i determines her preferences and model of inference, including:
  - her subjective belief about the relationship between gender and ability
  - her subjective belief about other evaluators preferences and beliefs

Model Setting

Evaluators

- A set of evaluators assess the worker's performance. For simplicity, assume that there is one evaluator per task, who reports evaluation $v_t \in \mathbb{R}$.
- Before evaluating task t, the evaluator observes:
  - the worker's gender g
  - evaluations on past tasks $h_t = (v_1, \ldots, v_{t-1})$, where $h_1 = \emptyset$
  - signal $s_t = q_t + \eta_t$ of quality of the current task, where $\eta_t \sim N(0, 1/\tau_\eta)$ is an independent random shock with precision $\tau_\eta > 0$
- An evaluator's type θ_i determines her preferences and model of inference, including:
  - her subjective belief about the relationship between gender and ability
  - her subjective belief about other evaluators preferences and beliefs
- The evaluator receives payoff $-\left(v - \left(q - c_g^i\right)\right)^2$ from reporting evaluation v on a task of quality q from a worker of gender g, where c_g^i is a type-specific taste parameter. Normalize $c_M^i = 0$.

Model Setting

Evaluators(continued)

Denition 1 (Preference-Based Partiality). An evaluator of type θ_i has a preference based partiality towards men if $c_F^i > 0$.

Definition 2 (Belief-Based Partiality). An evaluator of type θ_i has belief-based partiality towards men if $\hat{\mu}_M^i > \hat{\mu}_F^i$. This partiality is unbiased if $\hat{\mu}_M^i = \mu_M$ and $\hat{\mu}_F^i = \mu_F$ and otherwise is biased.

Discrimination

Let

$$D_i(h,s) \equiv v_i(h,s,M) - v_i(h,s,F) \tag{1}$$

denote the dierence between type θ_i's evaluation of a male and female worker conditional on observing history h and signal s, and let $D(h,s) \equiv E_\pi\left[D_i(h,s)\right]$ denote the expected dierence in evaluations across all types.

Denition 3 (Discrimination). A *woman faces* discrimination *from type* θ_i *at* (h,s) *if* $D_i(h,s) > 0$, *and faces* aggregate discrimination *if* $D(h,s) > 0$. A man faces (aggregate) discrimination if $D_i(h,s) < 0$ $(D(h,s) < 0)$.

Denition 4 (Discrimination Reversal). A discrimination reversal occurs at history h and signal s if there exists a history $h' \subset h$ such that women face discrimination at (h',s) and men face discrimination at (h,s).

Initial Discrimination

- Consider the evaluation of the first task from a worker of gender g by an evaluator who has subjective prior beliefs $(\hat{\mu}_F, \hat{\mu}_M)$ about average ability, preference parameter c_F, and observes signal s_1.
- The evaluator's prior belief about quality is normally distributed with mean $\hat{\mu}_g$ and precision $\tau_q \equiv \tau_a \tau_\varepsilon / (\tau_a + \tau_\varepsilon)$, i.e. $q_1 \sim N(\hat{\mu}_g, 1/\tau_q)$.
- Given the prior belief and signal distribution, the evaluator's posterior belief about quality conditional on observing s_1 is also normally distributed, $q_1 | s_1 \sim N\left(\frac{\tau_q \hat{\mu}_g + \tau_\eta s_1}{\tau_q + \tau_\eta}, \frac{1}{\tau_q + \tau_\eta}\right)$ .the optimal evaluation is equal to:

$$v(h_1, s_1, g) = \frac{\tau_q \hat{\mu}_g + \tau_\eta s_1}{\tau_q + \tau_\eta} - c_g \tag{2}$$

Initial discrimination is independent of the signal and equal to

$$D\left(h_1, s_1\right) = \left(\frac{\tau_q}{\tau_q + \tau_\eta}\right)\left(\hat{\mu}_M - \hat{\mu}_F\right) + c_F \tag{3}$$

Proposition 1 (Subjectivity of Judgement). If the evaluator has belief-based partiality, initial discrimination is decreasing in the precision of the signal τ_η and otherwise, initial discrimination is constant with respect to τ_η. As the signal becomes perfectly objective, $\tau_\eta \to \infty$, there is initial discrimination if and only if the evaluator has preference-based partiality.

Dynamics of Discrimination

Two cases:

1. evaluators share a common belief about the distribution of ability by gender, and this is common knowledge.
2. evaluators have heterogeneous beliefs some evaluators have belief-based partiality towards men and believe that all evaluators share the same beliefs, while other evaluators have no belief-based partiality but are aware that some evaluators do.

Dynamics of Discrimination

Suppose that all evaluators share a common prior belief about the distribution of ability by gender, have belief-based partiality, and this is common knowledge.

Proposition 2 (Impossibility of Reversal). Suppose there is a single type of evaluator with belief-based partiality and no preference-based partiality. Then fixing an evaluation history, discrimination decreases across periods but never reverses.

Dynamics of Discrimination

Possibility of Reversal in Misspecied Model

Proposition 3 (Possibility of Reversal). Suppose evaluators are the heuristic type θ_1 with probability $p \in (0,1)$ and the impartial type θ_2 with probability $1-p$.

1. For any initial evaluation v_1, there exist cut-offs $\overline{p} \in (0,1)$ and $\overline{s} \in \mathbb{R}$ such that for a low enough share of heuristic types $p \in (0, \overline{p})$ and a high enough second period signal, $s_2 > \overline{s}$, aggregate discrimination reverses in the second period, $D(v_1, s_2) < 0$.
2. For any second period signal s_2, there exist cut-offs $\overline{p}' \in (0,1)$ and $\overline{v} \in \mathbb{R}$ such that for a low enough share of heuristic types $p \in (0, \overline{p}')$ and a low enough ini- tial evaluation, $v_1 < \overline{v}$, aggregate discrimination reverses in the second period, $D(v_1, s_2) < 0$.

Dynamics of Discrimination

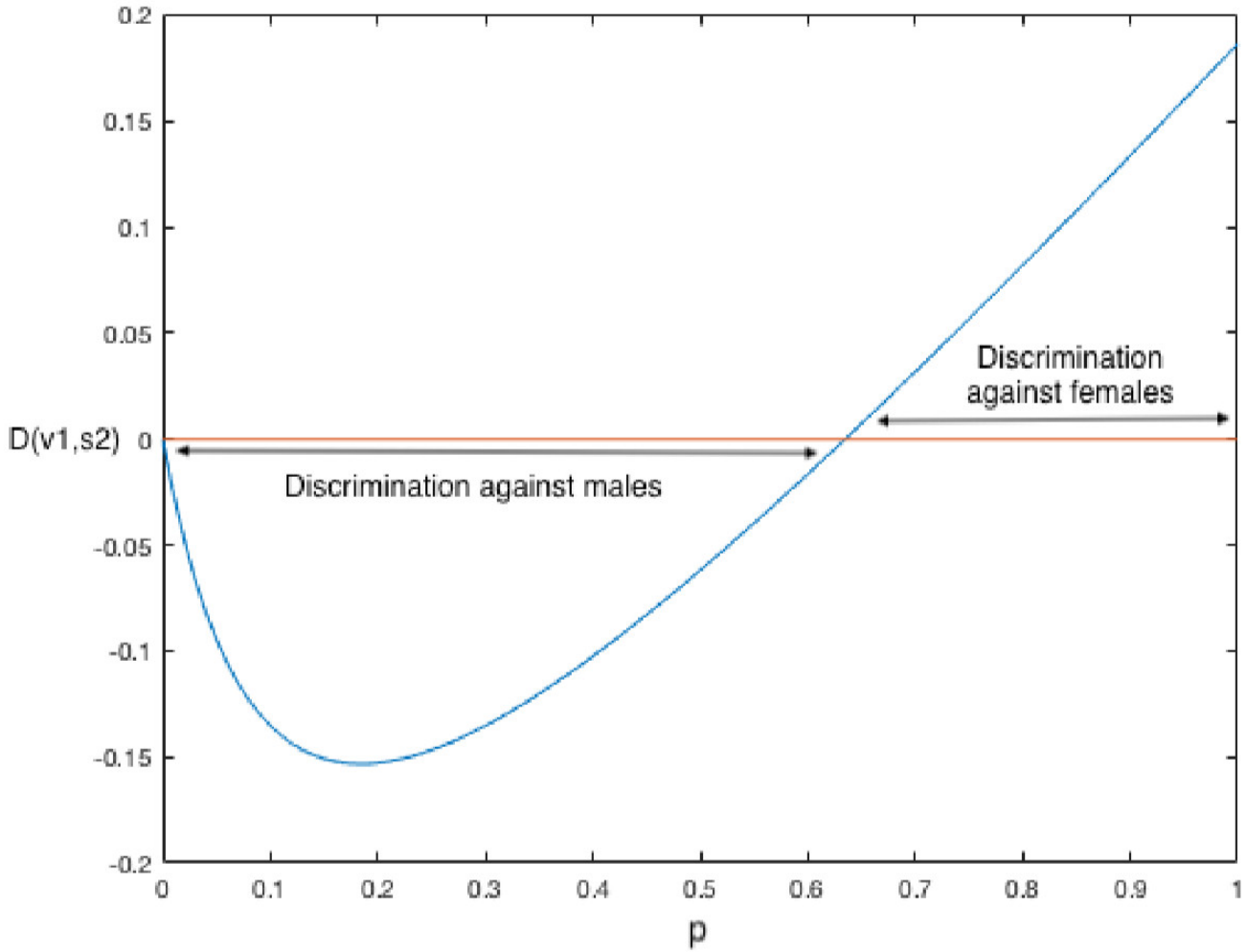


Figure 3: Second period discrimination, as a function of the proportion of heuristic evaluators.

Field Experiment

Experimental Design

Posting Questions

- 280 new accounts, with 140 male usernames and 140 female usernames. 140 were left as new accounts; these comprised the Novice accounts.
- For the other 140 accounts, we manually built-up the reputation of at least 100.(Advanced accounts)
- we wrote 280 novel mathematics questions ranging in level of diculty from upper-level undergraduate to early graduate.
- We measure discrimination as either the average change in reputation points per post (Rep) or the average number of upvotes net of downvotes per post (Net Votes).

Experimental Design

Posting Answers

- Original answers was generated to mathematics questions posted by other users on the forum,To examine how the subjectivity of judgment aects discrimination.
- Second set of 140 Novice accounts with no prior posts, split between 70 male usernames and 70 female usernames.
- research assistants worked in pairs('answerer' and 'poster')
- The standard of quality for answers is clear: determine whether or not the answer is correct. In contrast, there are multiple standards for judging the quality of a question.
- This difference in standards of quality should make judgment of questions more subjective than judgment of answers.

Experimental Results

Subjectivity of Judgment

Table 1: Effect of Gender on Evaluation of Novice Answers and Questions

	Answers Only		Questions Only		Answers & Questions	
	ΔRep	Net Votes	ΔRep	Net Votes	ΔRep	Net Votes
	(1)	(2)	(3)	(4)	(5)	(6)
Male	−1.38	−0.31	2.86	0.58	−1.38	−0.31
	(.97)	(.17)	(1.32)	(.27)	(1.16)	(.22)
Question					0.08	0.09
					(1.16)	(.22)
Male*Question					4.24	0.89
					(1.64)	(.32)
Constant	4.60	0.79	4.68	0.88	4.60	0.79
	(.69)	(.12)	(.93)	(.19)	(.96)	(0.16)
# Obs	135	135	135	135	270	270

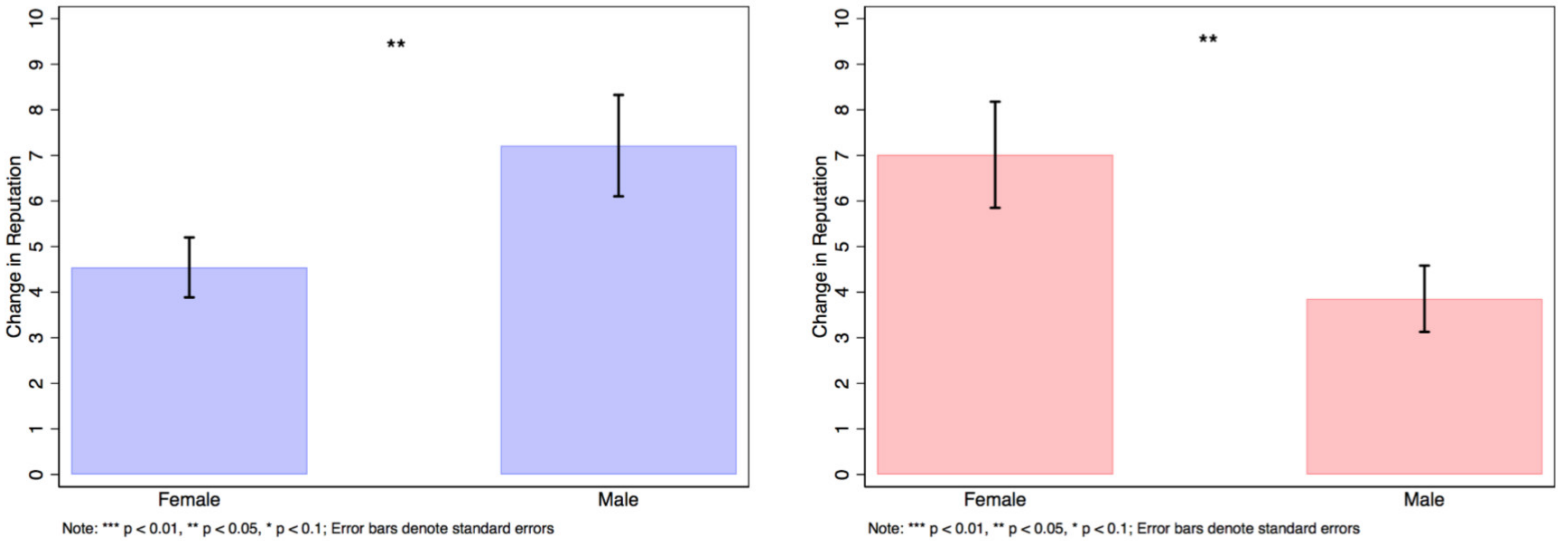
Experimental Results

Subjectivity of Judgment

- We find significant *initial discrimination* against females: Questions posted to accounts with female usernames accumulated signicantly fewer reputation points .
- There is a signicant mitigation of discrimination against female accounts for answers, relative to questions: the interaction effect between gender and type of post is positive and signicant in both specications.

Taken together, these results are inconsistent with discrimination due to preference based partiality. Rather, they support the theoretical prediction on how subjectivity affects discrimination when evaluators have *belief-based* partiality.

Experimental Results



(a) Novice Accounts

(b) Advanced Accounts

Figure 4: Average Reputation Difference for Questions

Experimental Results

Dynamics of Discrimination

Table 2: Eect of Gender on Evaluation Questions, Novice and Adva

	Advanced		Novice & Advanced		
	ΔRep	Net Votes	ΔRep	Net Votes	Binary
	(1)	(2)	(3)	(4)	(5)
Male	−3.16	−0.62	2.86	0.58	0.17
	(1.37)	(.28)	(1.36)	(.27)	(.08)
Advanced			2.33	0.49	0.09
			(1.35)	(.27)	(0.08)
Male*Advanced			−6.02	−1.20	−0.40
			(1.91)	(.38)	(.11)
Constant	7.01	1.38	4.68	0.88	0.56
	(0.97)	(.20)	(.96)	(0.19)	(.56)
# Obs	138	138	273	273	273

Experimental Results

Dynamics of Discrimination

- Questions posted to advanced female accounts accumulated more reputation points, ΔRep, than those posted to advanced male accounts:
  There is a dynamic *reversal of discrimination* between novice and advanced accounts.
- In our setting, not only is initial discrimination against females mitigated by reputation, but the direction of discrimination reverses, females are favored at higher reputations

Interpreting these findings through the lens of the theoretical framework, our results suggest that initial discrimination is driven by belief based partiality with bias.

Discussion and Conclusion

Three main results are:

1. significant gender discrimination exists at the initial stage, in the form of less reputation earned per post and fewer votes per post on questions posted by low reputation female accounts relative to questions posted by low reputation male accounts.
2. signicantly less gender discrimination at the initial stage for answers, where judgment of quality is less subjective relative to questions.
3. discrimination reverses for questions at more advanced stages, in that more reputation is earned and more votes are received on questions posted to high reputation female accounts relative to high reputation male accounts.

Thank You