Bias and Diversity II: Fairness at Equilibrium in the Labor Market

Lily Hu
Harvard University

October 16, 2017

These notes are based off a presentation by Lily Hu (Harvard University, PhD student in Applied Mathematics) for the section on "Bias and Diversity in the Mechanism Design for Social Good Reading Group and based on a paper-in-progress extending work presented at FATML 2017 [Hu and Chen, 2017]. Please contact the reading group organizers for any questions or comments.

1 Introduction

Empirically, racial inequalities in the U.S. labor market are incredibly well-documented and even more incredibly persistent over time. Black and white workers face divergent employment prospects and expected wage earnings upon entering the market, and this gap has even been shown to be widening over time. One of the most troubling aspects of this phenomenon is that this inequality has been concurrent with actual increases in minority worker protections, anti-discrimination law, and declines in at least explicit prejudicial attitudes in hiring. That is to say, that racial disparities have managed to persist and deepen within a realm of (more or less) equality of opportunity.

This paper takes a labor economics-influenced model of hiring/employment and casts the problem of discrimination as a dynamic game featuring strategic agents (workers, firms) who have incentives and expectations with a market with key endogenous features like costs, wages, and externalities.

Goal of the talk/paper:

1. Introduce a stylized model of labor market dynamics that can account for asymmetric group outcomes at steady-state due to rational best response strategies.

2. Install group symmetric outcomes labor market-wide at steady-state via a fairness intervention on firm hiring strategies.

3. Consider "welfare" properties of symmetric steady-state vs. asymmetric steady-state.

2 Labor Market Dynamics

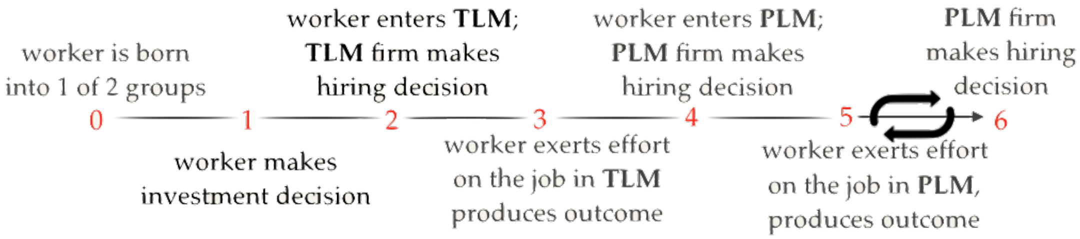
worker is born into 1 of 2 groups
0
1
worker makes investment decision
worker enters **TLM**; **TLM** firm makes hiring decision
2
3
worker exerts effort on the job in **TLM** produces outcome
worker enters **PLM**; **PLM** firm makes hiring decision
4
5
worker exerts effort on the job in **PLM**, produces outcome
PLM firm makes hiring decision
6

Figure 1: Timeline of labor market dynamics.

Model Overview:

- A worker belongs to one of two groups, each with collective reputations.
- Each group has the same distribution of worker ability levels ($F(\theta)$) but differing societal reputations π_μ.
- Labor market is segmented into Temporary Labor Market (TLM) and Permanent Labor Market (PLM).
- Cohorts of workers move through the labor market: enter TLM, pass into PLM after one firm interaction, cycle through firms/jobs in PLM, exit labor market.
- Workers hired by a firm earn a wage premium.

Agent actions:

- Workers: make pre-labor market human capital investment decisions, exert on-the-job effort.
- Firms: hire in the TLM based on workers' investment decisions, in the PLM based on workers' individual reputations.

Key Model Variable and Parameter Features:

- **Cost of investment** (c) is a function decreasing in worker ability level (θ), decreasing in group μ reputation (π_μ), and increasing in investment level ($\eta > 0$).
- **Cost of high effort exertion** (e) is a function of worker type (qualified vs. unqualified) and ability level (θ).
- **Wage** (w) is a function decreasing in proportion of workers producing good outcomes in the labor market (g_t).
- **Worker qualification status** (qualified vs. unqualified) is a hidden worker type; probabilistic function (γ) that maps to being qualified is increasing in investment level.

- **Probability of producing a good outcome** (p_H, p_Q, p_U) is higher for workers exerting high effort and higher for qualified workers exerting low effort than unqualified ones exerting low effort.

- **Individual reputation** (Π_i) is a summary statistic of the proportion of good outcomes (G) that a worker has had in her summary of past job performances (e.g. $\Pi_i(GBBGGB) = 0.5$).

- **Group reputation** (π_μ) is the proportion of all workers in group μ producing good outcomes in the labor market.

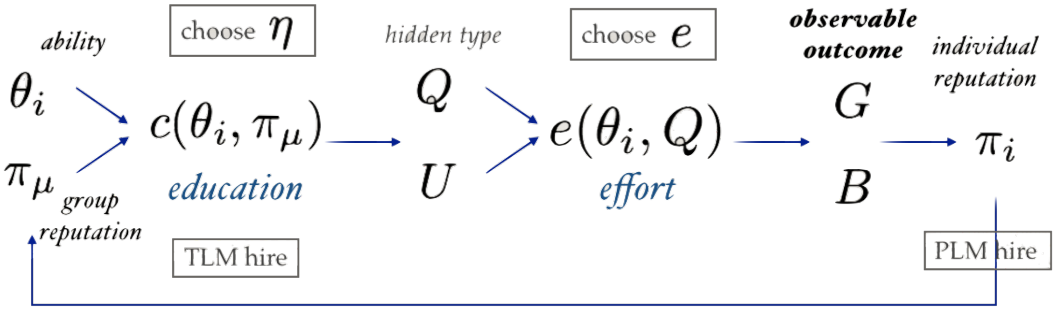

Figure 2: Flow of agent actions and attributes.

Figure 3: A worker i is born with an ability level θ_i and a group membership with reputation π_μ. Both of those factors contribute to her cost for education investment. She makes an investment level decision η, which begets a qualification status Q or U. She is granted entry into the TLM depending on the firm's hiring decision based on her investment. Her qualification type, which is hidden to firms now bears on her cost for effort exertion e as she enters the PLM and cycles through jobs. She stochastically produces G or B outcomes, yielding an individual reputation, which determines her future potential for being hired. Her own job performances also affect her group's reputation, which (with delay) affects future generations' investment costs.

Comment: Taste-based and statistical discrimination both share the quality that the discrimination is happening with respect to an observable trait, but statistical discrimination is different in that firms are only discriminating so as to be profit-maximizing, as opposed to having a taste for discrimination that would be built into their utility function.

Response: An argument against taste-based discrimination is that it should be able to be competed out by non-taste-based discriminators who can benefit from those who are irrationally discriminating against "perfectly good" workers. Another argument against modeling taste-based discrimination is that firms are no longer legally allowed to be paying different wages due to their group memberships for the same job, which is what effectively would have to happen under taste-based discrimination.

Question: Is there some distinction with respect to what your investment brings depending on what group you come from?

Answer: Regardless of group, if two workers invest at the exact same level (select the same η), they will have the exact same probability of being qualified for a particular job.

Question: Coming from a lower socioeconomic class, it's more costly for a person to make the same investment in terms of time and other resources; do you consider these differences in the model?

Answer: That is exactly what the distinct investment cost functions are. Groups have different cost functions; for two people of the exact same ability but of differing group memberships, the person from the group with lower social standing will face higher costs to achieve the same investment level.

2.1 Equilibrium Strategies

Firms want to hire workers who are most likely to produce good outcomes; workers want to maximize their expected reward (wage premium earned subtract costs). How will they play?

TLM strategies:

1. Firms, wanting to increase their probability of hiring a qualified worker of high ability, will set an investment threshold $\hat{\eta}$ (implicitly defined) on workers' investment levels, hiring all those with $\eta \geq \hat{\eta}$.

2. Workers, will invest at exactly $\hat{\eta}$ if they can afford to do so (cost is less than wage), at 0 if they cannot. Once in the TLM, they exert effort if they can afford to do so.

PLM strategies:

1. Since firms and workers contract in a repeated manner, firms seek self-enforcing contracts, leveraging workers' observable individual reputations, and set a reputation threshold $\hat{\Pi}^t = p_H - \Delta_t$ that is sensitive to workers' history lengths such that all workers with reputation $\Pi^t \geq \widehat{\Pi}^t$. ($\Delta_t$ is a monotonically decreasing sequence in t, cf. Rubinstein and Yaari [1983].)

2. Worker will exert effort whenever she can afford to do so.

Question: Are firms' hiring thresholds based on investment level or reputation level?

Answer: When workers are just entering the Temporary Labor Market, firms only see their investment level and group membership. When workers are cycling through jobs in the Permanent Labor Market, firms now see the additional information of the worker's individual reputation.

Question: The investment cost function is per individual, per group, per ability level?

Answer: Three parameters for each worker's investment cost: her choice of investment level (free), ability level (personal), and group reputation. So reputation is not hard-coded in, and reputations could flip, and what would matter is the ordering, not the explicit group membership.

3 Fairness Intervention

Statistical discrimination and group-blind hiring lead to asymmetric equilibria with group-asymmetric outcomes at steady-state—workers with the same ability but different group memberships make different investment choices. As a result, they face disparate wage and employment prospects. The lower reputation group is stuck with high costs that limit its members' capabilities to invest in human capital, which thus lowers their probability of being qualified. As a result, workers with high innate ability are inequitably barred from entering the skilled labor market. The asymmetry mirrors development bias, wherein members of a particular group face decreased opportunities for self-actualizing and realizing their potential, as opposed to reward bias, wherein members of a particular group are discriminated against in economic transactions on the explicit basis of group membership [Loury, 2009].

Since groups are fundamentally equal (same distribution of ability levels), on the basis of an anti-caste principle [Sunstein, 1994], workers should not face systematically disparate outcomes due only to social and historical factors. Thus a successful fairness intervention aims to achieve *group egalitarianism*, that is, symmetric group-wide outcomes at system steady-state.

Question: What is individual fairness? What is group fairness?

Answer: Individual and group fairness are two opposing "definitions" of fairness currently popular in the algorithmic fairness literature. For the sake of this particular conversation, I'm going to reduce individual fairness in the context of employment to being like "meritocratic fairness": in following the maxim "similar workers must be treated similarly" [Dwork et al., 2012], the axis of similarity that is relevant will be measuring workers' investment levels/likelihoods of being qualified. For this setup, individual fairness just guarantees that people with similar investment levels are hired with similar probability and thus face similar wage prospects. Group fairness, in contrast, is usually defined in reference to an idea of "fair" proportional representation of groups, usually along the lines of statistical or demographic parity. For example, statistical parity would require that in a world in which women constitute 50% of the population, women should constitute 50% of the selected candidates.

Question: Affirmative action is based more on holistic considerations, whereas what is proposed here is more of a quota-based system of hiring. Is there a relationship between these two forms of group-conscious hiring addressed here in the paper?

Answer: Arguably what's happening behind the scenes of lots of affirmative action-type policies currently in use is an informal quota system that isn't strictly legally allowed. For example, if you look at the demographic makeup of some college admissions, oftentimes it's going to be hovering around the same percentage makeup. And it's true that it's not a quota-based system; it's a holistic system, but there's some sort of operating belief about what the makeup of the particular incoming class population should be.

3.1 Proposed Fairness Intervention

Enforce a group-aware constraint $\mathcal{H}$ on firm hiring in the Temporary Labor Market such that for worker i of two potential group memberships W and B, $P(W) = P(W|\mathcal{H}(i))$ and

$P(B) = P(B|\mathcal{H}(i))$. In other words, under the hiring constraint, one's hiring status gives no more information about the group membership of the particular worker. This fairness constraint automatically entails *statistical parity*, wherein a group's population share must be reflected within the TLM pool of hired candidates. This fairness constraint in the TLM requires that firms set *group-specific* investment thresholds, which forces them to internalize the different costs facing workers of different groups. Firm hiring strategies in the Permanent Labor Market are not constrained in any way.

4 Steady-State Results

Result 1 (*Existence of Symmetric Steady-State*): The imposition of the proposed fairness intervention on firm hiring strategies in the TLM moves a system in which groups initially hold distinct societal reputations and thus face distinct costs for investment to eventually converge to a group-symmetric steady-state where workers do not face disparate wage prospects due to their group memberships.

Intuition: Forcing TLM hiring to abide by the fairness requirement prevents that entry from acting as a bottleneck that constrains the lower reputation group's representation in the labor market. It also ensures that groups retain identical ability distributions within the labor market. As a result, the only differences in group job performances stems from differences in the proportion of workers who are qualified within each group. This gap narrows with time as improved reputations feedback to improve cost conditions for future generations, thus increasing those workers' investment levels and thus their chances at being qualified, further improving the group's reputation, etc. Once group reputations converge, the investment costs for their members also converge, and thus, the TLM hiring constraint becomes obsolete, and firms naturally hire in manner consistent with statistical parity.

Result 2 (*Group-symmetric Pareto-dominates group-asymmetric*): Under particular market conditions, the group-symmetric steady-state outcome Pareto-dominates the asymmetric outcomes arising due to the popular unconstrained firm strategies of group-blind and statistical parity hiring.

Intuition: Reputation recovery is not possible under unconstrained hiring strategies, since the group with an initially lower reputation is inequitably barred from entry into the labor market, and thus cannot accumulate enough "good" outcomes from the proportion of its population that is hired within the market to lift the population-wide reputation. In such an asymmetric group outcome, workers in the lower reputation group who are high ability are unfairly blocked from the market where they otherwise would have been hired. Further, firms want to hire them! Under additional conditions, even those workers in the high reputation group who are only granted entry into the labor market in the unconstrained hiring regime are still not hired at equilibrium in the PLM anyway. Thus with no better off outcomes for workers in the high-reputation group and strictly better off outcomes for workers in the low-reputation group, the "fair" group-symmetric steady-state outcome Pareto-dominates.

Question: How about the welfare of the workers who are only granted entry into the labor market under the group-symmetric regime?

Answer: They're the ones who are strictly better off compared to the group-blind/statistical discriminatory case.

5 "Real world" connections and implications

There are indeed many moving parts to the model, but this many-stepped labor market sequence is common in the economics literature and the multiple parameters, types, choices/costs, etc. are well-founded and correspond roughly to stages and dynamics in the (very complicated) real world labor market.

Explanation of Modeling Choices

- Ability level (θ) vs. qualification status (Q/U)
  - A high ability worker is one who has the general attributes that bear on success in the realms of education and work, whereas a qualified worker is one who has the appropriate training and skills for a given job.
  - Very crudely, a worker is "born" with an ability level, while a worker "earns" a qualification status. In the model, a worker's ability level precedes her investment decision, which begets a qualification status.
- Education investment cost (c) vs. on-the-job effort exertion cost (e)
  - A worker's group reputation bears on her education investment cost, which reflects well-documented empirical findings that group membership affects one's access to material, social, and cultural resources (cf. Bowles et al. [2014] models effect of residential segregation on access to social capital). Concretely, being of a lower reputation group makes attaining the same level of investment more costly.
  - A worker's qualifications, or the extent to which her skill investment proved to be successful, becomes an overriding determinant once she is in the labor market, and thus on-the-job effort exertion is only a function of qualification and ability. But insofar as education investment bears on qualification status, it is clear that a worker's group membership continues to impact her labor market outcomes.
- Dual Labor Market Setup
  - Jobs in the Temporary Labor Market roughly correspond to short-term "internship" or trainee placements; while Permanent Labor Market positions are longer-term posts.
  - Cycling through jobs can be seen as moving between firms or changing roles/tasks within the same firm.

What's the upshot/prescription?: When hiring "new" workers or offering competitive internship or trainee slots, employ affirmative action policies. When hiring experienced workers and offering long-term posts, hire as you wish.

Question: Do you look at an intervention in which the cost of investment is subsidized such that hiring can subsequently be group-blind? Do you assess a comparison with that regime?

Answer: The short answer is that we do not consider that within the bounds of this paper. Someone who objects to this proposed intervention may say that the burden of ensuring fairness falls totally on the TLM firms, and that is true. So we may want to ask in future work whether there is a way to disperse the burden across many actors in the labor market or whether there is an external subsidy that can change the costs of investment in the beginning. The latter is something that the government is doing in programs like Head Start. We may want to think about what the convergence properties of those regimes are.

Question: Does this model address socioeconomic status inequalities? Or spill-over effects due to socioeconomic status improvement?

Answer: This model does not address that question, since the "next generation of members of one's group" is not a genetic understanding of the word "generation." It's simply cohorts of workers that belong to your group later in time, so there is no modeling of spillover effects of socioeconomic status improvement within a family. There is a paper by Bowles et al. [2014] that addresses human capital spillover effects in an agent's personal network.

6 Discussion

Although this work does not itself include much computational content, it is done with algorithmic contexts in mind especially considering the increasing usage of algorithmic agents used in employment settings to track employees' "reputations" and filter out/recommend candidates for job positions. One can develop the summary statistic form of the individual reputation model to be richer and include more data of a worker's history of performance outcomes and now ask a different but very closely related question about algorithmic fairness.

6.1 Remarks on "fairness" in algorithmic contexts

1. Modeling the problem: Dynamic game with strategic behavior vs. static machine classification task.

In nearly all societal domains in which fairness is an issue, past and current social relations and conditions differentially impact subjects and their sets of options, opportunities, resources, which mark their choices and outcomes today. So in many ways, the agent attributes seen today are not *a priori* givens, the way the standard learning theory formulation of the problem assumes. A dynamic model recognizes the powerful ripple effect of the past.

2. Addressing the root of inequalities: Development bias vs. reward bias

In proposing an intervention to combat bias and achieve fairness, it is paramount to disentangle the forms of bias that influence the system's dynamics and consider which bias the intervention is aimed at addressing. Much of the research about the sources of bias within a particular realm is an empirical one, but there is also much theoretical work to be done about the different interventions that should be undertaken to address different sources of bias. Within the labor market context, if one is concerned about development bias, which is what is considered here, the intervention should be different than one that would be proposed to address reward bias.

3. The timescale of "fairness": Long-term steady-state vs. short-term trade-offs

Conceiving of fairness at steady-state is a way to computationally formalize an ambitious long-term fairness project—group-egalitarianism—a worthy goal in itself that is difficult to conceive of when thinking about a short-term optimization problem trading off fairness with, say, immediate utility. Short-term fairness can prevent consideration of the long-term ramifications and externalities of actions such as group reputation effects. Also, the context of particular inequalities may call for different timescale conceptions of fairness.

6.2 Future work

Question: Are there quantitative measures for how temporarily "unfair" a treatment may be?

Answer: We can run some simulations and ask what the time to convergence looks like for some very generic function forms. This is the "How long will it be until affairs are fair?" question. We also may be interested in a regret-like notion, where you're not only concerned about how much time it takes to convergence but how much one is "losing" at each time point because of suboptimal hiring.

Question: What are other competing propositions for mitigating labor market inequalities?

Answer: Government subsidies to lower costs is one. Another is to soften the burden (maybe not hard constraint) in the TLM so it is not strictly quota-based, and then sprinkle some additional affirmative action-like policies in the PLM.

Question: Is it possible that one way to decrease the burden on the TLM hiring would be to not force total statistical parity, but to apply a particular amount of pressure in the direction of statistical parity, which will then push the system to equilibrium? Are there then effects on convergence time?

Answer: We haven't considered this explicitly. It could come down to the particular functional forms and parameterizations. In many other papers on collective reputation, there is usually a zone of group reputations, in which convergence is possible without a hard constraint, and then there are zones, in which once groups face such a large gap in reputations, there is no opportunity for reputation reclamation.

The intervention proposed here is not guaranteed to satisfy any popular notions of "fairness" along the path to the symmetric steady-state (individual, group, meritocratic, etc.), and there are surely losses of various kinds before reaching the fair outcomes. That is, we do pay a price for eventual fairness. Some related questions we may ask about the model for future consideration:

- What is the cost of such a fairness intervention in terms of time (convergence) and "lost productivity" (suboptimal hiring)?
- Are there ways to mitigate the "burden" of the hiring constraint on firms and still achieve group-equitable outcomes at steady-state?

References

Lily Hu and Yiling Chen. Fairness at equilibrium in the labor market. 2017.

Ariel Rubinstein and Menahem E Yaari. Repeated insurance contracts and moral hazard. *Journal of Economic Theory*, 30(1):74–97, 1983.

Glenn C Loury. *The anatomy of racial inequality*. Harvard University Press, 2009.

Cass R Sunstein. The anticaste principle. *Michigan Law Review*, 92(8):2410–2455, 1994.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226. ACM, 2012.

Samuel Bowles, Glenn C Loury, and Rajiv Sethi. Group inequality. *Journal of the European Economic Association*, 12(1):129–152, 2014.