MENTORING AND THE DYNAMICS OF AFFIRMATIVE ACTION*

Michèle Müller-Itten† and Aniko Öry‡

December 21, 2020

Abstract

We analyze the long-term workforce composition when the quality of mentoring available to majority and minority juniors depends on their representation in the workforce. A workforce with $\geq 50\%$ majority workers invariably converges to one where the majority is over-represented relative to the population. To maximize welfare, persistent interventions, such as group-specific fellowships, are often needed, and the optimal workforce may include minority workers of lower innate talent than the marginal majority worker. We discuss the role of mentorship determinants, talent dispersion, the scope of short-term interventions, various policy instruments and contrast our results to the classic fairness narrative.

JEL: D62, E24, I2, J15, J16, J24.

Keywords: Affirmative action, continuous time overlapping generations, human capital, labor participation, employment insecurity, mentoring, talent.

*We thank William Fuchs, Christiane Baumeister, Kirsten Cornelson, Thomas Gresik, Ben Hermalin, Sandile Hlatshwayo, Lisa Kahn, Maciej Kotowski, Mykhaylo Shkolnikov, Dan Silverman and two anonymous referees for insightful discussions and comments. We also thank Helena Laneuville Teixeira Garcia for excellent research assistance. Financial support from NSF (grant # SES 1227707) and the ISLA institute at the University of Notre Dame is gratefully acknowledged.

†University of Notre Dame. Email: michele.muller@nd.edu.

‡Yale University. Email: aniko.oery@yale.edu.

1 Introduction

Mentor relationships arise more readily between members of the same race, gender or other socioeconomic group. This can discourage young adults from selecting professions with few same-group mentors. As today's graduates turn into tomorrow's mentors, this effect persists and may exacerbate over time. The resulting labor force participation is suboptimal because young adults do not internalize the social benefit they create by becoming a future mentor. Affirmative action policies may reduce inefficiency, but what does the optimal workforce look like and how persistent does the policy need to be?

Motivated by this question, we provide a dynamic labor market framework to study inter-generational mentoring and its impact on labor force composition and total welfare. Our model builds on the following empirical findings: First, mentoring relationships are stronger between members of the same demographic group. Dreher and Cox Jr. (1996) find that female MBA students and MBA students of color are less likely to form mentoring partnerships with white men, and Ibarra (1992) finds differential patterns of network connectivity across genders. Second, the lack of similar role models affects the academic performance and labor market outcomes of minority students in ways that cannot be explained by differences in innate ability. This is why same-group teachers lead to a boost in student performance and graduation rates (Bettinger and Long, 2005; Dee, 2004, 2007; Fairlie et al., 2014). Third, these achievement differences arise early on and manifest themselves through different education choices. For instance, the undergraduate student body for economics has roughly the same composition as the academic workforce, indicating that the selection stems from education choices rather than differential attrition patterns (Bayer and Rouse, 2016).

We study labor force evolution when it is governed by workers' education decisions. The cost of education is dictated by idiosyncratic talent and the availability of mentors of the same group. However, juniors only account for the mentoring they receive, not the mentoring they provide for the next generation. Wages do not restore dynamic efficiency, because competition

and free entry prevents firms from internalizing the effect of today's hires on tomorrow's candidate pool. As a result, the long-run labor force composition can be inefficient.

With the following example, we highlight some driving forces in our general model and illustrate two of our key results: First, we show that the share of minority workers in the welfare-maximizing labor force can be higher than in the overall population. Second, we argue that the optimal intervention in such a situation is persistent.

Motivating example. A population is comprised of overlapping generations. Each generation consists of 80% majority group members ($i = 1$) and 20% minority group members ($i = 2$). One fourth of each group are of high talent (H), the remainder are of low talent (L). Each individual lives for two periods. In the first period, each individual (junior) may invest into costly education. In the second period, educated individuals (seniors) produce a surplus of one, which they receive as a wage. From the senior workforce, ten leaders are drawn at random and serve as mentors to currently enrolled juniors.

Low-talent individuals incur a private cost of c for education, high-talent individuals incur no cost. Each student also receives a mentoring payoff of 1 if at least one leader is from his own group.[1] However, the education decision is made before the identity of the mentors is revealed. As a result, a low-talent individual invests if and only if the expected mentoring boost of $1 - (1 - \phi_i)^{10}$ is large enough, or equivalently, when his own group makes up a large enough fraction ϕ_i of the senior labor force. We set the cost of education to $c = 2 - 0.65^{10} \approx 1.987$, so that in an unregulated market, juniors invest whenever their group makes up at least 35% of the senior workforce. High talent workers always invest in education, and indeed this is socially optimal.

We are interested in steady state workforce compositions where the pool of educated seniors equals the pool of students who invest. Table 1 reports the composition of four candidate workforces. Without intervention, compo-

[1] For simplicity, we assume that the mentoring benefit is realized even if the cost of education is zero.

Composition	(i)		(ii)		(iii)		(iv)	
	H	L	H	L	H	L	H	L
Majority participation	•	•	•		•	•	•	
Minority participation	•		•		•	•	•	•
% majority workers	94%		80%		80%		50%	
Total surplus	1.92		1.98		1.95		2.01	

Table 1: Motivating Example. The table reports the share of majority workers and the resulting surplus if the population segments indicated by • participate in the workforce.

sition (i) describes the only steady state where all investment decisions are individually rational. This composition excludes low-talent minority workers from the workforce. One could argue this is not a fair market since individual career outcomes favor the majority, even though there are no ex-ante talent differences across groups. Equity concerns may thus justify a policy intervention to reach a 'fair' composition such as (ii) or (iii). By either raising tuition for group 1 or lowering tuition for group 2, the policy maker can make these compositions individually rational. Perhaps surprisingly, welfare maximization motivates even starker interventions. Total welfare accounts for the mentoring externalities that workers exert on both low- and high-talent juniors; it is computed as the sum of aggregate output (or wage) plus any mentoring benefits net of education costs.[2] In this example, welfare is maximized in composition (iv), where the minority is over-represented in the workforce, resulting in a 50-50 split of the workforce at any point in time. To achieve this composition, the policy maker needs to modify the participation incentives of the majority through targeted interventions. ◇

Our general model also considers populations comprised of two groups and the cost of education as a function of innate talent and mentoring quality. However, we assume a continuous talent distribution and allow for other men-

[2] If there are H_i high- and L_i low-talent group-i workers, total welfare is given by aggregate output $\sum_{i=1}^{2} H_i + L_i$ plus mentoring benefits $\sum_{i=1}^{2}(H_i+L_i)(1-(1-\frac{H_i+L_i}{H_1+H_2+L_1+L_2})^{10})$ net of education costs $(L_1+L_2)c$.

torship functions that can be micro-founded in various ways.

The key parameters in our model are talent dispersion, mentor capacity and majority share. *Talent dispersion* measures the concentration of talent in the population. If all individuals have equal talent, the dispersion is zero. High-skill sectors (doctors, lawyers, professors) have high talent dispersion. We assume no ex-ante differences in talent distribution across the two groups. *Mentor capacity* captures the average number of mentees reached by a single mentor or, as in our initial example, the likelihood that a given senior becomes a mentor. This parameter is determined in practice by the type of mentor interaction: Capacity is high for classroom instruction but low for one-on-one coaching or when only a small fraction of seniors serve as role models or mentors. Finally, *majority share* refers to the percentage of majority group members in the overall population. The share is roughly 0.5 in the case of gender and larger in the case of race in the United States, where around 76.5% of the population is white.[3]

Investment in education is generally inefficient. A temporary intervention can move the economy from one steady state towards a more efficient one, as long as it is strong enough to affect convergence. However, if the magnitude of the intervention is too small, then the impact of a temporary intervention can be short-lived. For example, Bettinger and Long (2005) found that the positive effect of female instructors disappeared in some male-dominated fields. Similarly, Casas-Arce and Saiz (2015) show that political parties that were least affected by a gender-employment quota in Spain did not benefit in the long run. In order to identify the best temporary intervention, we thus compare total welfare generated at different steady states. We find that for sufficiently high mentor capacity or talent dispersion, the optimal stable steady state is such that top talent from both groups participate in the workforce. In sectors with high talent dispersion, the economy naturally converges toward this composition; but temporary affirmative action is warranted in sectors with low talent dispersion where mentor capacity is high and the initial workforce

[3]See U.S. Census Bureau QuickFacts, as retrieved from https://www.census.gov/quickfacts/fact/table/US/RHI125216 on 09/04/2019.

composition is nearly homogeneous. Undergraduate college education is such an example where classroom instruction allows for a high ratio of mentees per mentor, and learning relies on a broad set of skills. Thus, it may not be a coincidence that university admissions belong to the most visible Affirmative Action policies.

Yet, even the best steady state achieves less than maximal total welfare. We show that the optimal workforce over-represents the minority when the two population pools are of uneven size and mentor capacity is large. Consequently, a patient planner should persistently intervene in favor of the minority in many cases.

These policy implications are qualitatively different from those motivated by fairness, whose objective is 'equal opportunity for equal talent' regardless of group membership. This distinction is important because fairness was the main driver behind the initial affirmative action movement, and its vocabulary has since been adopted by the movement's opponents (Leonhardt, 2012). If fairness is the objective, affirmative action is primarily a remedy to historical injustice, and should render itself obsolete in a short time. Echoing this view, past discrimination takes center stage in the debate surrounding recent Supreme Court decisions on university admissions (Kahlenberg et al., 2014).

Persistent minority overrepresentation is not a 'fair' outcome: A minority student is in fact 'over-compensated' for his lack of suitable mentors relative to a majority student of equal talent. The crucial point of departure is that the equality of the two students is fictional under mentoring externalities: The minority student possesses mentoring skills that do more for future talent recruitment than those of his majority twin. A welfare-maximizing intervention remunerates him for that valuable skill. In particular, we show that the majority share in the population matters: Gender-based policies eventually become obsolete even under welfare maximization, but not necessarily those based on race or other minority characteristics.

Finally, we look at concrete policy instruments. We consider educational subsidies (scholarships), workplace hiring quotas, and mentor training. In our framework, the optimal educational subsidies are budget neutral in the long

run. Hiring quotas are equally effective only if the competitive environment allows for group-specific wages. However, when wage disparities are restricted due to cultural norms or firm-internal politics, hiring quotas cause significant crowding out of majority workers in the middle of the talent distribution. Because wages remain high, some majority workers keep investing in ex-post worth-less education and remain unemployed. This can result in strong opposition to hiring quotas among educated majority workers who are excluded from the labor market. To minimize this job insecurity, our model suggests that efficient wages under a hiring quota are *higher* for minority than for majority workers. Wage gaps that favor men are thus particularly harmful if they persist under hiring quotas, as is the case in Norway (Bertrand et al., 2014). Finally, we show that a nearly fair labor market emerges both as a stable steady state and as a close to optimal composition for large mentor capacity or when mentorship frictions disappear. Thus, the need for market intervention disappears if mentorship itself can be improved.

Our analysis is meant to be understood within a growing theoretical literature on workforce under-representation. The main takeaway from this literature is that different root causes of the observed hiring imbalance reach opposing verdicts on affirmative action: Under *taste-based discrimination* (Becker, 1957), affirmative action is essentially a zero-sum game where the benefit to the minority is offset by a direct utility loss of the majority, as documented by Besley et al. (2017) for political party leaders. Under *statistical discrimination*, employment quotas may actually reinforce negative stereotypes against certain groups (Coate and Loury, 1993; Fang and Moro, 2011; Fryer Jr, 2007). Indeed, when minority employment is mandated by law, firms may have to hire minority members even if they are unskilled. This in turn may actually *reduce* the minority's returns to education and thereby further lower equilibrium skill investment. Also, stereotypes are in many cases based on *biased beliefs* (Bohren et al., 2018; Bordalo et al., 2019). In the model of Bordalo et al. (2016), the bias stems from the representativeness heuristic (Kahneman and Tversky, 1972). If state-mandated diversity hires increase differences in the observed skill distribution between minority and majority workers, these

hires may exacerbate negative stereotypes. Finally, quotas are completely ineffective in altering beliefs when agents infer their personal success probability from their own group's employment history as in Chung (2000). We complement this discussion by showing that the benefit of affirmative action policies are understated and confounded if we ignore tangible mentoring complementarities.

Structurally, our analysis is in line with Ben-Porath (1967) who views human capital as being produced using innate talent and other inputs (which could be mentoring). Our paper builds on and extends the analysis of Athey et al. (2000), who study optimal promotion decisions in long-lived firms. We both assume that seniors offer an additive mentorship boost to juniors of varying talent, and that the size of this boost is increasing in the availability of same-group mentors. The crucial difference is that we do not assume that the two population pools are of equal size. This is crucial to obtain policy recommendations that go substantially beyond fairness concerns, and that require persistent intervention. Furthermore, unequal pools are arguably more suitable to capture demographic differences such as race or a "glass ceiling effect" in multilevel organizations, which affects optimal promotions (see p.25). Additionally, our results can be related to group identity norms as in Carvalho and Pradelski (2018). Contrary to Becker and Tomes (1979), Restuccia and Urrutia (2004) and Herskovic and Ramos (2017), we abstract away from income differences and focus instead on cultural and gender differences. Their analysis suggests that affirmative action is most effective if targeted towards the lower end of the income distribution.

The remainder of the paper is structured as follows: In Section 2 we set up our model of labor force participation and mentoring, and discuss a range of parametrizations that demonstrate the versatility of our framework. In Section 3, we analyze the steady states of an unregulated market and compare the labor force composition between temporary and persistent policy interventions. In Section 4, we contrast specific policy instruments and in Section 5, we discuss the robustness of our findings. Section 6 concludes by relating our analysis to the public discourse surrounding affirmative action.

2 Model

2.1 General model

We study an overlapping generations model with a unit mass of heterogeneous agents arriving in each period $t \in \mathbb{N}$. Each agent is indexed by a talent $x \in \mathbb{R}$ that is continuously distributed according to a cumulative distribution function F. Each agent belongs to either the majority group $i = 1$ with probability $b \geq 0.5$ or the minority group $i = 2$ with probability $1 - b$. Group membership is independent of talent, and we refer to b as the *majority share*. Hence, the mass of newly arriving agents with talent greater than x is equal to $b(1-F(x))$ for the majority group and $(1-b)(1-F(x))$ for the minority group.

Participation in the labor force is voluntary. Upon birth, each agent has the opportunity to elect an outside option with payoff zero. If an agent participates in the labor force, he lives for two periods and is called a junior in the first, and a senior in the second. As a junior, the agent pursues costly education. As a senior, the agent seeks employment and acts as a mentor for new juniors.

Juniors incur a cost of education, which consists of a fixed cost $c > 0$ that is reduced both by the junior's individual talent x and the group-specific strength of mentoring μ_i. In a period with mass L_i of group-i seniors and ℓ_i of group-i juniors, the mentorship boost μ_i for group-i juniors is determined by the *mentorship function* $\tilde{\mu}(L_i, L_{\neg i}, \ell_i, \ell_{\neg i}) \in [0, 1]$. We introduce generic structural assumptions in Section 2.2 and demonstrate the versatility of our framework with parametric examples in Section 2.3.

Seniors seek jobs in a competitive and unsaturated labor market. Earnings are determined through market forces: We assume that each unit mass of seniors contribute π units to a firm's profit flow. Assuming free entry of firms, the wage in an unregulated labor market is then equal to $w_i = \pi$ for both groups.

Individual rationality implies that a junior invests in education if and only if her expected lifetime earnings outweigh the cost of education. For a group-i junior with talent x who arrives in a period with $\boldsymbol{L} = (L_1, L_2)$ seniors of group 1 and 2, respectively, and $\boldsymbol{\ell} = (\ell_1, \ell_2)$ participating juniors of group 1 and 2,

respectively, education is individually rational if and only if

$$c - x - \tilde{\mu}\left(L_i, L_{\neg i}, \ell_i, \ell_{\neg i}\right) \leq w_i. \tag{IR}$$

As in Athey et al. (2000), we assume that there are no complementarities between talent and mentorship boost, and discuss the implications in Section 5.

Given a senior workforce $\boldsymbol{L}$, the junior workforce $\boldsymbol{\ell}$ corresponds to the mass of new arrivals for whom (IR) holds, i.e. the solution to

$$\begin{cases} \ell_1 = & b\left(1 - F(c - \tilde{\mu}(L_1, L_2, \ell_1, \ell_2) - w_1)\right) \\ \ell_2 = (1-b)\left(1 - F(c - \tilde{\mu}(L_2, L_1, \ell_2, \ell_1) - w_2)\right) \end{cases} \tag{1}$$

for $w_1 = w_2 = \pi$. In a regulated economy, wages are determined endogenously (see Section 4). As juniors turn into seniors, the dynamic system of labor force participation is characterized by $\boldsymbol{L}^{t+1} = \boldsymbol{\ell}^t$.

We are primarily interested in the group representation of a constant labor force, where $\boldsymbol{L}^t \equiv (\phi L, (1-\phi)L)$, and refer to $\phi \in [0,1]$ as the labor market *composition* and $L \in [0,1]$ as its total *size*. Without intervention, the labor force is constant only at a *steady state* where $\boldsymbol{\ell}^t = \boldsymbol{L}^t = \hat{\boldsymbol{L}} := (\hat{\phi}\hat{L}, (1-\hat{\phi})\hat{L})$ solves Equation (1). The steady state is *(Lyapunov) stable* if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \delta$, then $\|\boldsymbol{L}^t - \hat{\boldsymbol{L}}\| < \varepsilon$ for all $t > 0$.

We assume that scaling the entire population has no impact on mentorship quality, which implies that $\tilde{\mu}$ is homogeneous of degree zero,

$$\tilde{\mu}(L_i, L_{\neg i}, \ell_i, \ell_{\neg i}) \equiv \tilde{\mu}(kL_i, kL_{\neg i}, k\ell_i, k\ell_{\neg i}) \qquad \forall k > 0. \tag{M1}$$

To simplify notation, we then mostly use the one-dimensional restriction $\mu : [0,1] \to [0,1]$,

$$\mu(\phi_i) := \tilde{\mu}(\phi_i, 1-\phi_i, \phi_i, 1-\phi_i), \tag{2}$$

to describe the mentorship boost in any constant labor force with an own-group share $\phi_i = \frac{L_i}{L_i + L_{\neg i}} \in \{\phi, 1-\phi\}$, irrespective of total size $L = L_1 + L_2$.

A social planner can, in principle, induce any workforce composition and size, even if it is not a steady state. To determine the socially optimal work-

force composition, we maximize total surplus in a constant labor force. This welfare metric is relevant for a patient social planner who cannot adjust his diversity targets over time. Total surplus is measured per generation as total productivity net educational investments. A group-i worker of talent x generates an individual surplus of $\pi - c + x + \mu_i$ by participating, and zero otherwise. Integrating over all agents, the total surplus for a constant labor force of composition ϕ and size L yields

$$\begin{aligned} S(\phi, L) = \quad & b \int_{x \geq \hat{x}_1} \big(\pi - c + x + \mu(\phi)\big) dF(x) \\ & + (1-b) \int_{x \geq \hat{x}_2} \big(\pi - c + x + \mu(1-\phi)\big) dF(x) \end{aligned} \tag{3}$$

where $\hat{x}_1 = F^{-1}\big(1 - \frac{\phi}{b} L\big)$ and $\hat{x}_2 = F^{-1}\big(1 - \frac{1-\phi}{1-b} L\big)$ denote the marginal talent of group 1 and 2 workers, respectively. Perfect competition in the hiring market ensures that this surplus is entirely captured by educated juniors; their expected lifetime earnings outweigh their cost of education.

To position our findings within the policy debate around affirmative action, we formally define a *fair* labor market of constant workforce $(\phi L, (1-\phi)L)$ as one where no individual could be made better off by being born into the other group, i.e. $w_1 + \mu(\phi) = w_2 + \mu(1-\phi)$. The labor market is *biased towards the majority (minority)* if being born into this group is welfare enhancing, i.e. $w_1 + \mu(\phi) > (<)\, w_2 + \mu(1-\phi)$. We believe that this is how "fairness" is commonly understood in the public discourse. Finally, we say that a labor force is *dominated by the majority (minority)* if more than half of the labor force belongs to that group, $\phi > (<)\, 0.5$, and *over-represents the majority (minority)* when the share of workers belonging to that group is larger than the corresponding population share, $\phi > (<)\, b$. Note that whenever $b > 0.5$, a labor force can be dominated by the majority yet still over-represent the minority.

2.2 Mentorship and talent distribution

We now describe the general structural assumptions that we impose on the mentorship function and the talent distribution. We illustrate the versatility of the framework with two concrete parametric examples in Section 2.3.

Mentorship. The *mentorship function* $\tilde{\mu} : [0,1]^4 \to [0,1]$ describes the utility boost $\tilde{\mu}(L_i, L_{\neg i}, \ell_i, \ell_{\neg i})$ experienced by a group-i junior in a period with L_i ($L_{\neg i}$) own-group (opposite-group) seniors and ℓ_i ($\ell_{\neg i}$) own-group (opposite-group) juniors. We assume that $\tilde{\mu}$ is continuously differentiable, and that in any mixed labor force, an increase in the junior workforce weakly lowers the quality of mentorship, while an increase in own-group seniors strictly improves the quality of mentorship,

$$\frac{\partial \tilde{\mu}}{\partial \ell_i} \leq 0, \quad \frac{\partial \tilde{\mu}}{\partial \ell_{\neg i}} \leq 0, \quad \frac{\partial \tilde{\mu}}{\partial L_i} > 0 \quad \text{over } (0,1]^4. \tag{M2}$$

The sign of $\partial\tilde{\mu}/\partial L_{\neg i}$ can be positive, as would be expected when cross-group mentorship is effective. However, it could also be negative, when search frictions make it harder to find an own-group mentor in a senior workforce that is dominated by the opposite group. We only require that adding an own-group senior is weakly more beneficial than adding an opposite-group senior,

$$\frac{\partial \tilde{\mu}}{\partial L_i} \geq \frac{\partial \tilde{\mu}}{\partial L_{\neg i}}, \tag{M3}$$

and that adding an own-group junior lowers mentoring weakly more than an opposite-group junior,

$$\frac{\partial \tilde{\mu}}{\partial \ell_i} \leq \frac{\partial \tilde{\mu}}{\partial \ell_{\neg i}}. \tag{M4}$$

To determine how specific features of the mentorship function affect the labor force composition, it is useful to consider a family of mentorship functions $\{\tilde{\mu}_q\}$ identified by a single parameter $q > 0$ which we call the *mentor capacity*. The mentor capacity affects both the total and the marginal strength of mentorship. Letting μ_q denote the one-dimensional restriction of $\tilde{\mu}_q$ according to

Equation (2), we assume that the mentorship boost $\mu_q(\phi)$ in any mixed labor force $\phi \in (0,1)$ is pointwise increasing in q and satisfies the limits

$$\lim_{q\to 0} \mu_q(\phi) = 0, \quad \lim_{q\to\infty} \mu_q(\phi) = 1, \quad \text{and} \quad \lim_{q\to\infty} \mu_q'(\phi) = 0. \tag{M5}$$

In other words, high mentor capacity ensures a near-maximal mentoring boost at all representations $\phi \in (0,1)$. Letting

$$M_q(\phi) := \phi\mu_q(\phi) + (1-\phi)\mu_q(1-\phi) \tag{4}$$

denote the *total surplus generated by mentorship*, we assume that for any $\phi > 0.5$, there exists $Q_\phi \in \mathbb{R}$ such that

$$M_q'(\phi) < 0 \qquad \forall q \geq Q_\phi. \tag{M6}$$

This assumption plays a key role in resolving the tension between talent and mentoring. It ensures that talent is the deciding factor for large enough mentor capacity, and skews the composition that maximizes mentorship surplus $M_q(\phi)$ towards 0.5. We leave it to the parametric examples in the next section to convince the reader that the condition holds under a wide range of mentorship mechanisms.

Finally, to discuss stability, we need a technical condition that for all mentor capacities q the partial derivatives converge at comparable rates. Formally, we assume that for any $\delta > 0$, there exists a bound $K_\delta > 0$ such that

$$\|\nabla\tilde{\mu}_q(\phi, 1-\phi, \phi, 1-\phi)\|_\infty < K_\delta \mu_q'(\phi) \qquad \forall \phi \in (\delta, 1-\delta),\ \forall q > 0. \tag{M7}$$

This allows us to derive our main results by considering only the one-dimensional restriction μ rather than the full mentorship function $\tilde{\mu}$.

Talent. The cumulative talent distribution function F is continuously differentiable in x with full support over $(\underline{x}_F, \bar{x}_F)$, for some $\underline{x}_F, \bar{x}_F \in \mathbb{R} \cup \{\pm\infty\}$. For

realism and tractability, we assume that the range of talent is large enough,

$$\overline{x}_F > c - \pi - \mu_q(0.5) \qquad \text{and} \qquad \underline{x}_F < c - \pi - \bar{M}, \tag{F1}$$

where $\bar{M} = \max\left\{1, \sup_{\phi\in[0,1]} M_q(\phi) + (1-\phi)M_q'(\phi)\right\} < \infty$ is determined by the mentorship function. The first condition ensures that the most talented juniors have positive individual surplus in a constant labor force with an equal mass of either group. This is necessary for positive labor supply in any mixed steady state. The second condition ensures that the least talented workers have negative individual surplus under any mentoring boost (since $\mu(\phi) \leq 1$ always) and their participation also lowers social surplus. This simplifies some of the proofs because we do not have to worry about corner solutions, while no substantive insights are lost.

To determine the role of talent on the labor force composition, it is useful to consider a family of talent distributions $\{F_\lambda\}$ identified by a single parameter $\lambda > 0$ which we call the *talent dispersion*. Talent dispersion measures the spread of talent in the population, and we assume that the support of F_λ weakly increases in the set-inclusion sense, with $\lim_{\lambda\to\infty} \bar{x}_{F_\lambda} = \infty$. Consequently, whenever Property (F1) holds for some λ and q, it also holds for all larger $\lambda' \geq \lambda$ or $q' \geq q$, by monotonicity of $\hbar x_{F_\lambda}$ and $\mu_q(0.5)$. Further, for any x that is (eventually) inside the support of F_λ, we assume that the hazard rate converges to zero pointwise,

$$\lim_{\lambda\to\infty} \frac{F_\lambda'(x)}{1 - F_\lambda(x)} = 0. \tag{F2}$$

Loosely speaking, this ensures that the upper tail grows fast relative to the mass of agents with talent near x.

2.3 Parametric examples

There are various sensible assumptions regarding the mechanisms that determine the strength of mentoring, or regarding the talent distribution in the population. In this section, we present a range of parametric examples that

fit our general framework:

Mentorship. Our first example considers matching frictions, that are made explicit in a discrete matching market.[4] We consider random bipartite network between $n(L_1 + L_2)$ seniors and $n(\ell_1 + \ell_2)$ juniors. Links are drawn independently, and represent successful mentoring relationships. Members of the same group i are linked with probability p_{ii}, members of opposite groups with a lower probability $p_{ij} < p_{ii}$. Specifically, each mentor is assigned to q mentees on average. Parameter q is high for industries where the relevant skills are imparted through classroom instruction, and low where mentoring requires individual coaching.[5] Out of all mentees, a fraction $s \in [0, 1)$ is drawn from the pool of same-group juniors while the remainder is drawn from the entire junior population (partial assortativity). A within-group mentor assignment is always successful, while an across-group mentor assignment only with probability $\delta \in [0, 1)$ (homophily). Putting these together, the total probability[6] for a link between a group-i senior and a group-j junior is equal to

$$p_{ij} = \begin{cases} \frac{sq}{n\ell_i} + \frac{(1-s)q}{n(\ell_1+\ell_2)} & \text{if } i = j \\ \delta \frac{(1-s)q}{n(\ell_1+\ell_2)} & \text{otherwise.} \end{cases}$$

[4]There is a vast empirical literature on the strong positive effects of same-group role models and mentors. Notably, the performance gap between white and underrepresented minority students drops by 20-50 percent in courses taught by a minority instructor (Fairlie et al., 2014), and one year with an own-race instructor increases math and reading scores by 2 to 4 percentile points (Dee, 2004). These performance boosts are especially pronounced for minority students of the highest ability levels (Carrell et al., 2010; Ellison and Swanson, 2009). The literature also documents a bias where faculty fails to identify talented minority students (Card and Giuliano, 2016), bases track recommendation on gender stereotypes (Carlana, 2019) or perceives other-race students as inattentive (Dee, 2005). Similar positive effects arise when schools are segregated by gender (Jackson, 2016), suggesting that peer effects may amplify such patterns. Fully assortative matching would solve this problem but is rarely feasible, particularly in professions with high degrees of specialization or geographic fragmentation, where mentor assignment is primarily dictated by expertise or location. Moreover, if mentoring occurs in groups, assortativity reduces benefits of peer group diversity and can raise segregation concerns.

[5]Decreasing trends in the time invested in mentoring (DeLong et al., 2008) will also affect market dynamics through this channel.

[6]For n large enough, the link probabilities are non-degenerate.

A junior enjoys a mentorship boost of 1 if and only if he or she is in at least one successful mentoring relationship. In a very stark fashion, this captures the idea that there are decreasing returns to scale from mentorship for an individual junior.

By the Law of Rare Events, the number of mentors per junior can be approximated by the Poisson distribution as n grows. The success probability of finding a mentor, and hence the expected mentorship boost for a junior of group i, converges to

$$\tilde{\mu}(L_i, L_{\neg i}, \ell_i, \ell_{\neg i}) = 1 - e^{-sq\frac{L_i}{\ell_i} - (1-s)q\frac{L_i + \delta L_j}{\ell_1 + \ell_2}}. \tag{5}$$

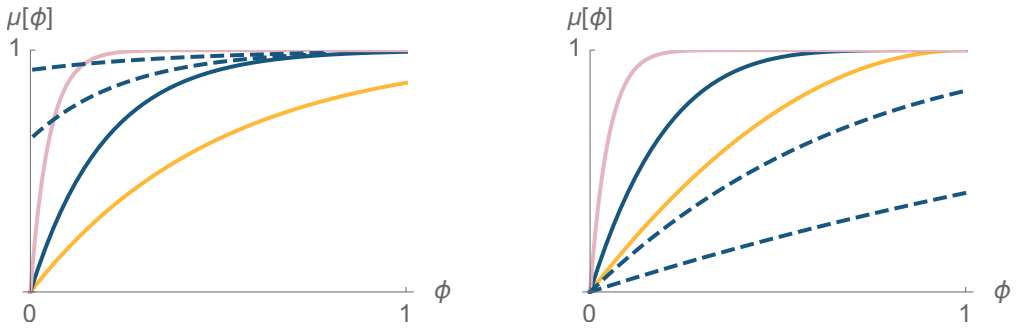
Figure 1a shows how increasing mentor capacity ($q \uparrow$), improving across-group mentoring ($\delta \uparrow$), or increasing assortativity ($s \uparrow$) impacts the shape of the mentorship function $\tilde{\mu}$. Note that (M4) is satisfied since assortativity is imperfect ($s < 1$), and (M3) is satisfied since group-match matters ($\delta < 1$).

Our second example expands upon the example in the introduction, and views mentors primarily as role models.[7] Formally, we assume that in every generation q leaders are randomly appointed, and serve as role models for juniors. The number of group-i role models is thus distributed according to a Binomial distribution with success probability $L_i/(L_1 + L_2)$. A junior with $k \in \mathbb{N}_0$ same-group role models enjoys a mentorship boost of $1 - \delta^k$ for some $\delta \in [0, 1)$. This incorporates a more subtle version of decreasing returns from mentoring, and approaches the binary version from above when $\delta \to 0$. The expected mentorship boost to a group-i junior is then equal to

$$\tilde{\mu}(L_i, L_{\neg i}, l_i, l_{\neg i}) = 1 - \left(\frac{\delta L_i + L_{\neg i}}{L_i + L_{\neg i}}\right)^q. \tag{6}$$

Figure 1b shows how increasing the number of leaders ($q \uparrow$) or weakening the decreasing returns ($\delta \uparrow$) impacts the shape of the mentorship function $\tilde{\mu}$. Note

[7] Role models are important as documented for example in Kofoed et al. (2019) who find that cadets are more likely to pick their officer's branch if they have the same gender or race. Similarly, Porter and Serra (2019) find that female economics students who are exposed to female role models are more likely to choose economics as their major.



(a) random assignment of mentors (b) random appointment of role models

Figure 1: Parametric mentorship functions. Solid lines indicate mentor functions for $q = 2$ (yellow), $q = 5$ (blue) and $q = 20$ (pink) with $\delta = s = 0$. Increasing δ or s lowers the slope and curvature of the mentoring function as indicated by the dashed lines.

that (M3) is satisfied because returns from additional mentors are decreasing ($\delta < 1$).

We share the term 'role models' with Chung (2000), but we reach different conclusions because of different assumptions on the benefits that juniors reap. In Chung (2000)'s model, the only information learned from a role model is whether "someone like me" can succeed in a world that is static. In temporarily lowering the hurdles for one group, affirmative action makes their success less meaningful to post-policy juniors. As also acknowledged by the author, we believe this is a narrow view of what role models do. They teach us not only *if*, but also *how* "someone like me" can succeed. There is actual, group-specific knowledge created when people from one's own group get to participate in the labor force. Second, the value of role models goes beyond mere information. Role models are also trail blazers, who transform the working culture into one that is more welcoming for successors like them.

Talent. In terms of the talent distribution, our results are driven by the talent dispersion λ. High talent dispersion means that a small elite possesses abundant talent, low dispersion reduces this heterogeneity. In applications, λ is particularly large for specialized education that requires rare skills, such as

for doctors, lawyers or actors.

Formally, we only require the conditions discussed in Section 2.2. Practically, we are mainly interested in families of distribution where the mean talent is constant while the variance increases with λ. Good candidates are the normal distribution with any fixed mean and variance λ, the gamma distribution with fixed mean and scale parameter λ or the uniform distribution over $(-\lambda, \lambda)$. However, the results also apply when dispersion increases along with the mean. For instance, talent could be determined by the sum of k different skills, each exponentially distributed with mean λ, which would then yield a Gamma distribution with rate parameter k and shape λ.

The most striking difference between these families lies in the range of possible talent. In the normal distribution, talent is unbounded both above and below. The Gamma distribution imposes a lower bound on talent, and as λ increases, the likelihood of near-zero talent increases. Finally, the uniform distribution imposes both an upper and lower bound on talent, but the range is increasing in λ.

3 Optimal Labor Force Composition

In this section, we determine the labor force composition that emerges in an unregulated steady state, and then show that the composition is generally suboptimal. Intuitively, mentoring complementarities generate a tension between talent recruitment and mentoring efficiency: Only a homogeneous labor force ($\phi \in \{0, 1\}$) ensures perfect within-group mentor assignments, but a mixed labor force ($0 < \phi < 1$) harnesses the top talent from both groups.

3.1 Steady States of an unregulated economy

In an unregulated economy, each worker earns his marginal product, $w_1 = w_2 = \pi$. The break-even talent in a steady state of composition $\hat{\phi}$ is given by

$$\hat{x}_1(\phi) := c - \pi - \mu(\phi) \qquad \text{and} \qquad \hat{x}_2(\phi) := c - \pi - \mu(1 - \phi)$$

for each group i. Property (F1) ensures that both $\hat{x}_1(\phi)$ and $\hat{x}_2(\phi)$ are strictly above $\underline{x}_F$, and at least one of them is strictly below $\bar{x}_F$ in a steady state. This avoids corner solutions where all members of one group participate, or where no one from either group participates. The assumptions do not rule out *homogeneous* steady states where only one group participates, i.e. $\hat{\phi} = 0$ with $\hat{x}_1(0) > \bar{x}_F$ or $\hat{\phi} = 1$ with $\hat{x}_2(1) > \bar{x}_F$.

Our first result identifies necessary and sufficient conditions for each type of steady state, and shows that mixed stable steady state are generally biased towards the majority. We say that the steady states of an economy tend toward a finite subset $\Phi \subset [0,1]$ as a parameter tends to infinity if and only if for every $\delta > 0$, all steady states $(\hat{\phi}, \hat{L})$ satisfy $\min_{\phi' \in \Phi} |\hat{\phi} - \phi'| < \delta$ for sufficiently large parameter values.

Proposition 1 (Steady States)**.** *Consider an economy that satisfies Property (F1).*

(a) The economy admits two homogeneous steady states $\hat{\phi} \in \{0,1\}$ if and only if the most able individuals require some mentorship boost to participate,

$$c - \bar{x}_F - \mu(0) \geq \pi. \tag{hSS}$$

The homogeneous steady states are stable whenever the inequality is strict.

(b) The economy always admits a mixed steady state $\hat{\phi} \in (0,1)$.

(c) If $b > 0.5$, any majority-dominant workforce converges to a steady state that is biased towards the majority.

(d) As mentor capacity $q \to \infty$, the economy admits a stable steady state near b. The steady states of the economy tend towards $\{0, b, 1\}$.

(e) As talent dispersion $\lambda \to \infty$, the economy admits a stable steady state near b. The steady states of the economy tend towards $\{b\}$.

Proof. See Appendix A.2.1. □

We find that when mentoring is *required* for investment of even the most educated individuals (hSS), a stable homogeneous steady state exists (claim a). This is precisely because group-i investment ceases once the group is severely underrepresented in the workforce.

Every economy also admits at least one mixed steady state, though it may be unstable (claim b). An economy may admit multiple stable steady states, including some where the workforce is dominated by the population minority, $\hat{\phi} < 0.5$. South Africa is an example that readily comes to mind, where 80% of the economically active population is Black African, yet they still hold only 14.3% of top management jobs even over 20 years after the end of apartheid (BBC News, 2019). Our analysis suggests that mentorship disparities can sustain such a bias towards the minority indefinitely. Achieving a fairer market may require active government intervention.

Whenever the majority initially dominates the workforce, however, mentorship frictions invariably push the economy towards a bias in favor of the majority (claim c). First, we show that a majority-dominant senior workforce always attracts more juniors of the majority than of the minority. Still, that in itself does not constitute a bias. Recall that a labor market is fair if the break-even talent for each group is identical. Because talent is equally distributed, a fair steady-state labor market with $\hat{x}_1(\phi) = \hat{x}_2(\phi)$ would therefore mirror the composition of the population, $\hat{\phi} = b$. A bias towards the majority means not just that the majority will always dominate the workforce ($\hat{\phi} > 0.5$), but that it will eventually be over-represented ($\hat{\phi} > b$). We show this by ruling out any steady states with composition $\hat{\phi} \in (0.5, b]$. To relate this to the example above, it means that minority dominance never emerges spontaneously, but is always the result of active interventions that institute minority rule either by forcing an initial composition $\phi < 0.5$ or artificially skewing the payoffs in favor of the minority group.

Finally, if either mentor capacity or talent concentration is large, labor supply hardly responds to differences in mentor availability. If q is large, this is because even a small representation yields a near-maximal mentorship boost; if λ is large, this is because there are very few juniors in the middle of the talent

distribution who could be swayed to participate with mentorship. Thus, for any interior ϕ, the ratio of group-1 to group-2 individuals with talent above $\hat{x}_i(\phi)$ converges to $b : 1 - b$, ruling out any other steady states.

3.2 Welfare-maximizing steady state

In a first step, we provide policy recommendations for interventions that are limited in time. A patient social planner would then redirect the economy towards the steady state that maximizes surplus. We denote this composition by ϕ^*_{SS}.

Proposition 2 (Optimal Steady State)**.** *For sufficiently large mentor capacity q or high talent dispersion λ, the surplus-maximizing (stable) steady state is nearly fair, $\phi^*_{SS} \approx b$.*

Proof. See Appendix A.2.2. □

As mentor capacity increases, even a handful of minority mentors can provide a near-perfect boost to minority juniors. As a result, the efficiency tension resolves in favor of talent recruitment, and surplus is maximized at a nearly fair steady state. Highly concentrated talent makes all other workforce compositions unsteady, and so that part of the result follows directly from Proposition 1(e). In other words, temporary market intervention is warranted when minority participation threatens to vanish in an industry where mentoring is sufficiently broad and differences in talent are not very pronounced. This makes sectors with low specialization and mentoring through classroom instruction (for example undergraduate education) prime candidates for temporary course correction in favor of the underrepresented group.

Still, temporary intervention does not achieve a workforce that accurately reflects the diversity in the population. Although the mixed steady states tend towards fairness, the minority remains *under*represented at the mixed steady state in the sense that $\hat{\phi} > b$ for any finite q or λ by Proposition 1(c). This is because minority mentors are harder to come by, making it impossible to sustain proportional participation without ongoing intervention.

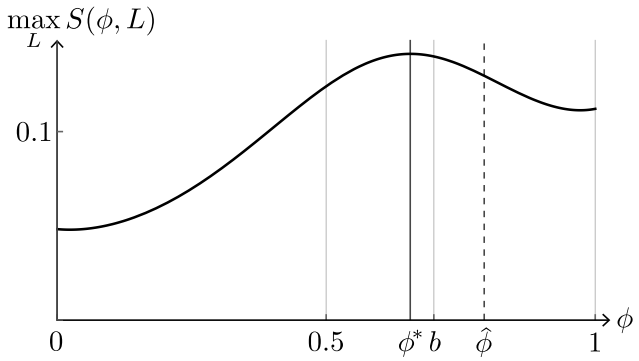


Figure 2: Social surplus as a function of labor force composition ϕ.

Figures are obtained using mentorship function (6), uniform talent distribution $U(-1, 1)$, and parameter values $b = 0.7$, $c = 2.2$, $\pi = 1$, $\delta = 0$ and $q = 5$.

3.3 Optimal long-run intervention

Perhaps surprisingly, we now show that an optimal long-term policy often over-represents the minority. In doing so, we are agnostic about the exact implementation of the policy goal. We simply assume that the planner can dictate each individual participation decision, which is equivalent to choosing the marginal talent of participants in either group. In Section 4, we show that the optimal policy can be implemented through educational scholarships or hiring quotas.

Before stating the proposition, it is useful to visualize the surplus function defined in (3). Figure 2 depicts the maximal surplus across compositions. In this example, the optimal labor composition is biased in favor of the minority ($\phi^* < b$), but the steady state composition is biased in favor of the majority ($\hat{\phi} > b$). The surplus generated under the optimal composition ϕ^* is 9% higher than the steady-state surplus. The figure also illustrates that a fair labor market (with composition b) achieves a near-optimal surplus, and may be easier to implement for political reasons. The shape of the surplus function in this example is typical. For small mentor capacity, the optimum is found at the right boundary ($\phi^* = 1$). For large mentor capacity, the optimum is majority dominant but biased in favor of the minority. The next proposition formalizes these claims.

Proposition 3 (Optimal Intervention)**.** *The optimal labor force composition* ϕ^* *depends on mentor capacity* q *as follows:*

(a) If $c - \pi > \bar{x}_F$ *and* q *is sufficiently small, a homogeneous labor force is optimal.*

(b) If $b > 0.5$*, the optimal labor force is always dominated by the majority,* $\phi^* > 0.5$*. However, as long as mentor capacity is sufficiently large,* $q > Q$*, the optimal labor force is biased in favor of the minority,* $\phi^* \in (0.5, b)$*.*

(c) The optimal composition converges to that of the population $\lim_{q\to\infty} \phi^* = b$*.*

For large enough talent dispersion λ*, the surplus-maximizing economy is biased towards the minority (majority) whenever* $M_q'(b) < 0$ *(*$M_q'(b) > 0$*).*

Proof. See Appendix A.2.2. □

Figure 3 illustrates the different regions described in the proposition by plotting the optimal composition ϕ^* as a function of the mentor capacity q. We start by observing that whenever there is a population majority, the optimal labor force is always dominated by that group (see claim b). In the South Africa example mentioned above, this would motivate an intervention in favor of the dominated majority.

When the maximal talent is not too high and mentor capacity is small, we then show that the most efficient labor force excludes the minority (claim a, region A in Figure 3). Including even just the most talented minority members dilutes mentoring for the majority, and this effect can outweigh for small mentor capacities. Note however, that a homogeneous labor force is never optimal when the upper bound on talent, $\bar{x}_F$, is large enough.

Larger mentor capacities make the mentoring dilution less costly for the majority. At some point, this implies that the optimal labor market is actually biased in favor of the minority (claim b, region B in Figure 3), though the size of the bias disappears in the limit (claim c). In such a market, the policy maker recruits minority workers with talent *below* the marginal majority worker – not just as a transitory course correction, but as an ongoing policy. The stark result

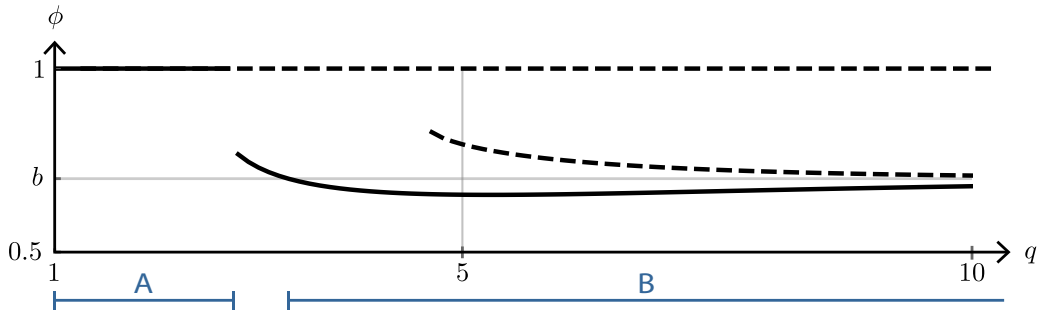


Figure 3: Optimal composition as a function of mentor capacity q (solid line). Stable steady state compositions are indicated with dashed lines.
Figure is obtained using mentorship function (6), uniform talent distribution $U[-1, 1]$, and parameter values $b = 0.7$, $c = 2.2$, $\pi = 1$, and $\delta = 0$.

has a simple intuition: Workers don't internalize their own positive mentoring externality on future generations. When mentors are efficient (q large), the social returns warrant minority subsidies that exceed the inherent mentoring advantage of the majority.

A critical expression that emerges in the proof is whether $M'_q(b)$ is negative or positive. This expression can be rewritten as

$$\mu(b) + b\mu'(b) < \mu(1-b) + (1-b)\mu'(1-b), \tag{7}$$

which compares the mentoring gain from marginal over-representation between the two groups. An additional majority participant enjoys a mentorship boost of $\mu(b)$ and improves the available mentoring for his entire group of mass b by $\mu'(b)$. When the marginal returns from mentoring are sufficiently decreasing, it is more beneficial to instead add an additional minority participant, since $\mu'(1-b) \gg \mu'(b)$. As long as talent is sufficiently dispersed, this is sufficient to ensure that a minority bias is optimal (claim d).

Since a composition $\phi \in (0.5, b]$ never emerges in a steady state, Propositions 1 and 3 jointly imply that a sufficiently patient planner intervenes persistently in favor of the minority in industries where mentoring has sufficiently decreasing returns-to-scale, and individual surplus is primarily driven by talent rather than mentoring. In particular, there is no reason to assume

that affirmative action policies render themselves obsolete by virtue of their own success. This is contrary to the 2003 Supreme Court ruling which argued that "race-conscious admissions policies must be limited in time" and expected them to disappear within 25 years.[8]

Proposition 3 also points to differences between race- and gender-based affirmative action. Since both genders are equally prominent in the population ($b \approx 0.5$), we expect gender-based policies to be necessary only in the short run, but see grounds for ongoing race-based policies (since $b \gg 0.5$). In other words: Extrapolating from a model with equal population pools (Athey et al., 2000) might lead us to believe that course corrections are not necessary in industries where skill recruitment dominates mentoring – when in fact these are the precise situations where surplus maximization requires ongoing intervention.

It is also useful to contrast our results with Athey et al. (2000)'s conjecture regarding the "glass ceiling effect," referring to the well-known phenomenon that group-imbalance increases in higher echelons of the career ladder.[9] In their model, senior management plays the role of a surplus-maximizing social planner. The authors observe that for $b = 0.5$, a marginal population increase of one group shifts the optimal labor force composition towards that new majority. From that, they conjecture that (a) a population increase for one group shifts the optimal bias towards this group, and (b) representation inequalities are exacerbated at each level in an organizational hierarchy (Athey et al., 2000, p.778f). Our analysis warrants a more nuanced view: (a) While a population increase shifts the optimal workforce representation towards that group, the *bias* may actually be in favor of the other group, and (b) faced with an uneven middle management, optimal promotion decisions at the top may over-represent the dominated group.[10] Thus, mentoring frictions alone do not provide a persuasive rationale for increasing attrition across echelons of the career ladder, at least not if promotion decisions are surplus maximizing.

[8] Grutter v. Bollinger, 539 U.S. 306 (2003), pages 309-310.

[9] Matsa and Miller (2011) report that women only make up 6% of corporate CEO's and top executives, despite representing 47% of the labor force.

[10] Moreover, if promotions represent the top end of the talent distribution, there is no reason to assume any interactions between levels at all.

4 Policy Instruments

We now turn our focus to the practical implementation of a policy that modifies labor participation. While the previous section determined the socially optimal workforce $\boldsymbol{L}^* = (\phi^* L^*, (1-\phi^*)L^*)$ assuming direct control over the talent cutoffs $x_1^* = F^{-1}\left(1 - \frac{\phi^*}{b} L^*\right)$ and $x_2^* = F^{-1}\left(1 - \frac{1-\phi^*}{1-b} L^*\right)$, here we ask how the policy maker can implement cutoffs (x_1^*, x_2^*) using available policy tools. We compare three methods that can be expressed within our simple model: group-specific tuition, hiring quotas and mentor training.

Educational incentives. The most direct market intervention modifies the cost-benefit analysis of prospective students through a combination of group-specific fellowships and tuition hikes. Let $\boldsymbol{\Delta} \in \mathbb{R}^2$ denote such a transfer schedule where Δ_i represents the net transfer to individuals in group i. These transfers are assumed to be available to *all* interested minority students. It is straightforward that ability-based fellowships *only* affect the extensive margin if the available pool exceeds the unregulated student supply.[11] Because the labor market remains unrestricted, expected returns to education remain equal to $w = \pi$. Equation (1) ensures that participation $\boldsymbol{L}^*$ is individually rational, given a status quo labor force (L_1, L_2), if and only if

$$\Delta_i = c - \pi - \tilde{\mu}(L_i, L_{\neg i}, L_i^*, L_{\neg i}^*) - x_i^* \qquad \forall i = 1, 2.$$

After one period of intervention, the status quo labor force becomes $\boldsymbol{L}^*$, but the policy needs to stay in effect since $\boldsymbol{L}^*$ is generally not a steady state.

We now show that once the surplus-maximizing mixed labor force is reached, it can be maintained in a way that is budget-balanced. Since there are ϕ^* transfers of Δ_1 for every $1-\phi^*$ transfers of Δ_2, budget balance at $\boldsymbol{L}^*$ requires that

$$0 = \phi^* \Delta_1 + (1-\phi^*)\Delta_2. \tag{8}$$

[11] This may explain why studies such as Prenovitz et al. (2016) fail to observe additional minority recruitment for competitive scholarship programs on a limited budget.

Consider now a change in the constant labor force by a marginal increase in total size. As long as the labor force composition is maintained, the individual surplus of any participating workers is unaffected, but the increase adds $\phi^* dL$ group-1 workers with individual surplus $-\Delta_1$ and $(1-\phi^*)dL$ group-2 workers with individual surplus $-\Delta_2$. Since L^* is chosen optimally, the total effect must be zero, which implies Equation (8).

Labor Force Quotas. Alternatively, the policy maker can restrict the recruitment decisions of firms by setting caps on the group composition of new hires. Norway is a prime example of such an approach, since it was the first country to mandate quotas for managerial boards in publicly listed companies – a sector with high skill concentration. Spain and Iceland have since implemented similar policies (Egan, 2012). Politicians typically distinguish between so-called hiring "goals" and more explicit "quotas," but that distinction is largely semantic from an economic perspective (Fryer and Loury, 2005). For that reason, we simply impose upper limits on the proportion of majority group members among all educated new hires.[12] We call a quota ϕ^* *binding* at $\boldsymbol{L}$ if it forces the firm to recruit more minority members than they would otherwise. Formally, if $\boldsymbol{\ell}$ denotes the solution to Equation (1) under wages $w_i \equiv \pi$, ϕ^* is binding if and only if $\phi^* < \frac{\ell_1}{\ell_1+\ell_2}$.

With a quota, the policy maker controls only the composition of the market, while market forces determine the size of the labor force. We study two cases, depending on whether the market allows for wage differentials based on minority membership. We need some new notation since regulation may jeopardize employment security: We denote the mass of *educated* and *employed* individuals by $\overline{\boldsymbol{\ell}} \geq \boldsymbol{\ell}$ respectively. We assume that all educated group members are equally likely to get hired since firms care only about productivity, so that the expected earnings under wages w_i are equal to $\ell_i/\bar{\ell}_i \cdot w_i$.

When firms can choose wages freely, any oversupply of educated group-i workers would drive wage w_i to zero. As long as the mass of workers willing to

[12] Only quotas with restrictions on education can be effective. Otherwise, firms could always costlessly meet any quota by hiring unqualified minority workers at a wage of zero.

work at zero wage is small enough, this implies that all educated workers find employment, $\bar{\boldsymbol{\ell}} = \boldsymbol{\ell} = (\phi^*\ell, (1-\phi^*)\ell)$.[13] Under a binding quota and given a status quo labor force (L_1, L_2), the size of the cohort ℓ and the market wages w_i are then uniquely determined by the market clearing equations

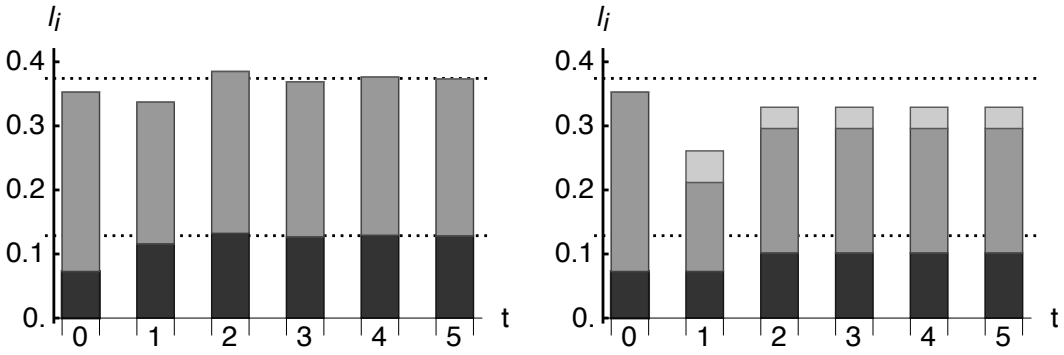
$$\begin{cases} \phi^*\ell = & b\,(1 - F(c - \tilde{\mu}(L_1, L_2, \phi^*\ell, (1-\phi^*)\ell) - w_1)) \\ (1-\phi^*)\ell = & (1-b)\,(1 - F(c - \tilde{\mu}(L_2, L_1, (1-\phi^*)\ell, \phi^*\ell) - w_2)) \\ \pi = & \phi^* w_1 + (1-\phi^*) w_2. \end{cases}$$

The first two expressions restate the individual rationality constraints (1). The third equation is the zero-profit condition for firms with the required share of minority workers. Note that it is equal to budget balance (8) when wage bonuses are restated as subsidies $\boldsymbol{\Delta} = \boldsymbol{\pi} - \boldsymbol{w}$. This implies that a binding quota *raises* minority and *depresses* majority earnings relative to the unconstrained market, $w_1 < \pi < w_2$. Our model does not distinguish between monetary wages and other job perks; similar effects are obtained if firms offer benefits that are geared towards the minority. It also implies that the social planner and the myopic firms agree on the optimal labor size at the surplus-maximizing composition. It may, however, take several generations until the labor size approaches the optimal level, as illustrated by Figure 4a.

In some industries, social or legal pressure prohibits paying unequal wage to employees in the same position, $w_1 = w_2$.[14] The zero-profit condition forces these market wages to π. However, a binding quota caps the demand for group-1 workers at $\ell_1 = \frac{\phi^*}{1-\phi^*}\ell_2$, while all educated minority workers are hired, $\ell_2 = \bar{\ell}_2$. Workers factor this employment insecurity into their participation

[13] If there are many majority workers who obtain an education regardless of earnings, market wages are an insufficient instrument to guide participation decisions. We omit the formal conditions because we do not think that these offer additional insight into realistic scenarios.

[14] This is the stated rationale behind the presidential memorandum 'Advancing Pay Equality Through Compensation Data Collection' (Presidential Memorandum, 79 Fed.Reg. 20751 (Nov.04, 2014), www.federalregister.gov/d/2014-08448). Firms also have internal incentives to avoid group-specific wages, as pay gaps can have detrimental effects on worker morale and firm output if the gaps are not easily accounted for by productivity differences (Breza et al., 2017).



(a) group-specific wages (b) common wages

Figure 4: Labor force evolution starting at the mixed steady state for $t = 0$, under a quota that imposes the welfare-maximizing composition. For each generation, the bars show the mass of minority (black), employed majority (gray) and unemployed majority participants (light gray). Dotted lines indicate the welfare-maximizing workforce.
Figures are obtained using mentorship function (6), uniform talent distribution $U[-1, 1]$, and parameter values $b = 0.7$, $c = 2.2$, $\pi = 1$, $\delta = 0$ and $q = 5$.

decision, transforming the individual rationality constraints (1) into

$$\begin{cases} \bar{\ell}_1 = & b\,(1 - F(c - \tilde{\mu}(L_1, L_2, \bar{\ell}_1, \bar{\ell}_2) - \frac{\phi^*}{1-\phi^*}\frac{\bar{\ell}_2}{\bar{\ell}_1}\pi)) \\ \bar{\ell}_2 = & (1-b)\,(1 - F(c - \tilde{\mu}(L_2, L_1, \bar{\ell}_2, \bar{\ell}_1) - \pi)). \end{cases}$$

Job insecurity for the majority is the only driver for the change in participation rates, and so this must occur in equilibrium, as illustrated in Figure 4b. In practice, this means that majority workers waste their own resources on an ex-post worthless education and dilute mentoring efficiency for everybody else. Of course, such a feature greatly reduces the appeal of workplace quotas in situations where wage is sticky or subject to social scrutiny.

Mentor training. For large mentor capacity q, the tension between talent recruitment and mentorship efficiency disappears. A nearly fair labor market emerges both as a stable steady state (Proposition 1), and this composition is close to optimal (Proposition 3). Thus, the need for market intervention dis-

appears if mentorship itself can be improved through cross-group exposure,[15] mentor training, and networking support for minority youth.[16] Our model can be helpful in highlighting the benefits of increased mentor capacity, but estimating the cost and feasibility of such improvements is mainly an empirical question.

5 Robustness

Determinants of productivity. We currently assume that productivity is binary and affected only by schooling. One natural extension is to also allow for innate talent and mentoring to affect productivity directly, so that education brings productivity of a worker with talent x and mentorship boost μ by $\pi_1 x + \pi_2 \mu + \pi_3$. If firms can observe talent and vary wages by worker, then their perfect competition ensures that workers still reap their entire individual surplus. A change of parameters $\tilde{c} = \frac{1}{1+\pi_2} c$, $\tilde{\pi} = \frac{1}{1+\pi_2} \pi_3$ and $\tilde{x} = \frac{1+\pi_1}{1+\pi_2} x$ can then map this situation into our existing model. Indeed, the mapping transforms individual surplus

$$\underbrace{\pi_1 x + \pi_2 \mu + \pi_3}_{\text{productivity}} - \underbrace{(c - x - \mu)}_{\text{cost}} = (1 + \pi_2)(\tilde{\pi} - (\tilde{c} - \tilde{x} - \mu))$$

into a constant multiple of the individual surplus in our standard model with the new parameters. Since unregulated dynamics and social surplus are both governed by the sign and relative size of individual surplus, the qualitative results of our paper carry over unchanged.

[15] Dobbin and Kalev (2016) show that programs that increase contact among groups (in particular formal mentorship programs or voluntary task forces) are most effective in affecting the minority representation among managers. Similarly, Beaman et al. (2009) show that increased exposure to female leaders (through a quota system) reduces biases.

[16] One of the main goals of the presidential initiative "My Brother's Keeper" is to connect young men of color to mentoring and support networks (Obama, Barack. "Remarks by the President on 'My Brother's Keeper' Initiative." *The White House*, Office of the Press Secretary, 27 Feb 2014, https://obamawhitehouse.archives.gov/the-press-office/2014/02/27/remarks-president-my-brothers-keeper-initiative).

Non-additive surplus. One limitation of our approach is the assumption that talent and mentoring affect surplus additively. One can imagine scenarios where the effectiveness of a given mentoring relationship depends not only on the mentor's group membership or education, but is affected (positively or negatively) by mentor or mentee talent, and the mentor's experience as both a mentor and a mentee.

We share this assumption with Athey et al. (2000), but it is difficult to relax. Mathematically, the main difficulty is the history-dependence in mentoring and talent, which complicates the steady state analysis. It is, however, possible to qualitatively anticipate the impact of non-additive surplus under interventions that maintain a constant labor force. In these interventions, both the distribution of educated talent and mentor experience are fixed in the long term. When the minority is over-represented $0.5 < \phi < b$, the conditional talent distribution among educated majority workers is left-censored, relative to that of minority workers, and a typical majority student experiences better mentoring. As such, over-representation is reinforced if low-talent students have greater returns from mentorship, if there is a negative correlation between individual talent and mentoring skill, or if poorly mentored students turn into more "attuned" mentors later in life. The opposite is true if high-talent students are more receptive mentees or if high-talent/well-mentored workers are more resourceful mentors.

Uneven talent distribution. While we firmly believe that only a model with equal talent distribution across groups can inform optimal policy, some situations call for a "conditionally optimal" policy, given the planner's constraints. For instance, a university may not be able to address systemic differences in access to primary education and may be confronted with sizeable test gaps across applicants from two groups. Our analysis also has relevance for the optimal admission policy in these settings. In particular, assume that the talent distribution F_1 of the majority first order stochastically dominates that of the minority, F_2. Whenever $L_1 \geq L_2$, the majority has an advantage both mentoring-wise and talent-wise. In an unregulated economy, this

would inevitably lead to a bias in favor of the majority analogous to our result from Proposition 1(c). Yet, letting $\phi = \frac{b(1-F_1(\hat{x}))}{b(1-F_1(\hat{x}))+(1-b)(1-F_2(\hat{x}))} \geq b$ denote the share of majority workers that arise from a fair labor force with $\hat{x}_1 = \hat{x}_2 = \hat{x}$, note that Property (M6) implies that for large enough mentor capacity q, the marginal benefits from an additional minority mentor outweighs those from an additional majority mentor. In turn, this implies that a bias in favor of the minority is optimal, which requires persistent intervention. Thus, while the minority may not be over-represented in the conditionally optimal workforce, the bias would still be in favor of the minority, and achieving this bias would still require ongoing intervention.

6 Conclusion

We do not want this paper to be read in isolation. Affirmative action has many important consequences and we focus primarily on its interaction with mentoring and its impact on workforce composition. However, we believe that awareness of the surplus consequences of mentoring complementarities is crucial for the public discussion. On the most basic level, the insights of our model are these: People differ in their ability to recruit and mentor top talent from different socio-demographic backgrounds. Often, mentors are most effective within their own social group. Like any other skill set, it makes sense to remunerate group-specific mentoring ability according to the shortness of its supply and its impact on future surplus. However, such remuneration does not arise in an unregulated economy due to firm competition because minority workers do not account for their future positive externalities in their education decisions. Affirmative action policies, in the form of scholarships or hiring quotas, can act as a correcting force. To guide the design of the optimal policy, a keen understanding of wage determination is necessary to avoid unintended consequences.

Our main contribution is to show that the scale of these externalities can be far larger than previous models suggest, to the point where they warrant an on-going subsidy towards the minority that goes beyond a correction of

historical under-representation. The optimal remuneration often generates a target workforce that is *more diverse* than the population, where the net cost of education is *lower* for the minority than for the majority. More specifically, this arises in sectors that require rare skills, and when the marginal mentorship gains from increased representation are larger for the minority than for the majority. We consider concrete policy instruments to achieve the optimal workforce composition. We argue in favor of widely available minority scholarships over hiring quotas, and encourage strategies that improve mentorship and connectivity for minority workers.

A Additional Proofs

A.1 Mathematical Lemmata

This first lemma serves to simplify notation in subsequent proofs.

Lemma 1. *Let* $\boldsymbol{m}(\phi) := \frac{1}{\mu'(\phi)}\nabla\tilde{\mu}(\phi, 1-\phi, \phi, 1-\phi)$. *Assumptions (M1) to (M4) imply that the following hold for all* $\phi \in (0,1)$ *and* $L > 0$*:*

$$\nabla\tilde{\mu}(\phi L, (1-\phi)L, \phi L, (1-\phi)L) = \frac{\mu'(\phi)}{L}\boldsymbol{m}(\phi) \tag{9}$$

$$m_1(\phi) + m_3(\phi) = 1 - \phi \tag{10}$$

$$m_2(\phi) + m_4(\phi) = -\phi, \tag{11}$$

$$m_1(\phi) > 0 \geq m_4(\phi) \geq m_3(\phi) \tag{12}$$

$$m_1(\phi)m_3(1-\phi) \leq m_2(\phi)m_4(1-\phi) \tag{13}$$

$$m_1(\phi)m_1(1-\phi) \geq m_2(\phi)m_2(1-\phi). \tag{14}$$

Lastly, for any $\delta > 0$*, there exists* $K_\delta \in \mathbb{R}$ *such that* $\|\boldsymbol{m}(\phi)\|_\infty < K_\delta$ *for all* q *and all* $\phi \in (\delta, 1-\delta)$.

Proof. The first conditions follows by homogeneity of degree zero (M1):

$$\nabla\tilde{\mu}(\phi L, (1-\phi)L, \phi L, (1-\phi)L) \overset{\text{(M1)}}{=} \frac{1}{L}\nabla\tilde{\mu}(\phi, 1-\phi, \phi, 1-\phi) = \frac{\mu'(\phi)}{L}\boldsymbol{m}(\phi).$$

The next two properties follow by homogeneity of degree zero (M1) and the fact that μ is continuously differentiable:

$$\begin{aligned} m_1(\phi) + m_3(\phi) &= \lim_{\Delta \to 0} \frac{\tilde{\mu}(\phi+\Delta, 1-\phi, \phi+\Delta, 1-\phi) - \mu(\phi)}{\Delta \cdot \mu'(\phi)} \\ &\overset{(M1)}{=} \lim_{\Delta \to 0} \frac{\mu\left(\frac{\phi+\Delta}{1+\Delta}\right) - \mu(\phi)}{\Delta \cdot \mu'(\phi)} = \lim_{\Delta \to 0} \frac{\frac{\phi+\Delta}{1+\Delta} - \phi}{\Delta} = 1-\phi, \\ m_2(\phi) + m_4(\phi) &= \lim_{\Delta \to 0} \frac{\tilde{\mu}(\phi, 1-\phi+\Delta, \phi, 1-\phi+\Delta) - \mu(\phi)}{\Delta \cdot \mu'(\phi)} \\ &\overset{(M1)}{=} \lim_{\Delta \to 0} \frac{\mu\left(\frac{\phi}{1+\Delta}\right) - \mu(\phi)}{\Delta \cdot \mu'(\phi)} = \lim_{\Delta \to 0} \frac{\frac{\phi}{1+\Delta} - \phi}{\Delta} = -\phi. \end{aligned}$$

The sign conditions Property (M2) imply that m_1 is positive and m_3, m_4 are negative, and Property (M4) imposes $m_3 \leq m_4$.

To show Equation (13), note first that the left side is always weakly negative. Thus, the inequality is automatically satisfied whenever $m_2(\phi) \leq 0$. Otherwise, the result follows from multiplying the two inequalities (M3) and (M4).

Next, Property (M3) imposes $m_1 \geq m_2$ which – as long as both $m_2(\phi)$ and $m_2(1-\phi)$ are positive – directly implies condition (14). When $m_2(\phi)$ and $m_2(1-\phi)$ have opposite signs, inequality (14) holds simply because the left side is positive and the right side is negative. Finally, when both $m_2(\phi)$ and $m_2(1-\phi)$ are negative, their product is maximal when they both achieve their lower bound

$$m_2(\phi) \overset{(11)}{=} -\phi - m_4(\phi) \overset{(12)}{\geq} -\phi \quad \text{and} \quad m_2(1-\phi) \overset{(11)}{=} -(1-\phi) - m_4(1-\phi) \overset{(12)}{\geq} -(1-\phi),$$

implying $m_2(\phi)m_2(1-\phi) \leq \phi(1-\phi)$. Since $m_1 > 0$, the left side is minimal at the lower bound

$$m_1(\phi) \overset{(10)}{=} 1-\phi-m_3(\phi) \overset{(12)}{\geq} 1-\phi \qquad \text{and} \qquad m_1(1-\phi) \overset{(10)}{=} \phi - m_3(1-\phi) \overset{(12)}{\geq} \phi,$$

hence $m_1(\phi)m_1(1-\phi) \geq \phi(1-\phi) \geq m_2(\phi)m_2(1-\phi)$.

Lastly, the boundedness of $\boldsymbol{m}$ stems directly from the boundedness assumption in Property (M7). □

Lemma 2. *For any $x, y \in (\lim_{\lambda\to\infty} \underline{x}_{F_\lambda}, \infty)$, the talent distribution satisfies*

$$\lim_{\lambda\to\infty} \frac{1-F_\lambda(x)}{1-F_\lambda(y)} = 1 \tag{15}$$

$$\lim_{\lambda\to\infty} \int_y^{\bar{x}_{F_\lambda}} \frac{1-F_\lambda(x)}{1-F_\lambda(y)} dx = \infty. \tag{16}$$

Proof. To show Equation (15), note first that it is without loss of generality to assume that $x \geq y$. By Property (F2), we can then write

$$\begin{aligned} 1 \geq \frac{1-F_\lambda(x)}{1-F_\lambda(y)} &= 1 - \frac{F_\lambda(x)-F_\lambda(y)}{1-F_\lambda(y)} = 1 - \int_y^x \frac{F'_\lambda(t)}{1-F_\lambda(y)} dt \\ &\geq 1 - \int_y^x \frac{F'_\lambda(t)}{1-F_\lambda(t)} dt \xrightarrow{\lambda\to\infty} 1 + 0. \end{aligned}$$

Furthermore, consider any $M > 0$. Since $\bar{x}_{F_\lambda} \to \infty$, there exists Λ_1 such that $\tilde{x} := \bar{x}_{F_{\Lambda_1}} > 2M + y$. Furthermore, there exists Λ_2 such that $\frac{1-F_\lambda(\tilde{x})}{1-F_\lambda(y)} > 1/2$ for all $\lambda > \Lambda_2$. Hence,

$$\int_y^{\bar{x}_{F_\lambda}} \frac{1-F_\lambda(x)}{1-F_\lambda(y)} dx \geq (\tilde{x}-y)\frac{1}{2} > M \qquad \forall \lambda > \max\{\Lambda_1, \Lambda_2\},$$

which implies Equation (16). □

A.2 Relegated proofs

A.2.1 Steady States

We can rewrite Equation (1) as

$$\mathbf{0} = \boldsymbol{G}(\boldsymbol{L}^t, \boldsymbol{L}^{t+1}) := \begin{bmatrix} L_1^{t+1} - b\big(1 - F\big(c - 1 - \tilde{\mu}(L_1^t, L_2^t, L_1^{t+1}, L_2^{t+1})\big)\big) \\ L_2^{t+1} - (1-b)\big(1 - F\big(c - 1 - \tilde{\mu}(L_2^t, L_1^t, L_2^{t+1}, L_1^{t+1})\big)\big) \end{bmatrix} \tag{17}$$

which fully describes the evolution of the dynamic system as long as $L_1^t, L_2^t > 0$. Recall that a steady state $\hat{\boldsymbol{L}}$ is Lyapunov stable if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \delta$, then $\|\boldsymbol{L}^t - \hat{\boldsymbol{L}}\| < \varepsilon$ for all $t > 0$. To determine whether a steady state $\hat{\boldsymbol{L}}$ is stable, we will consider the linearization

of the system in order to derive a sufficient condition for a steady state to be stable and a sufficient condition for a steady state not to be stable. To simplify notation, we write the Jacobian as

$$X(\hat{\boldsymbol{L}}) := \left.\frac{\partial \boldsymbol{L}^{t+1}}{\partial \boldsymbol{L}^{t}}\right|_{\hat{\boldsymbol{L}}} = -\left[\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{L}^{t+1}}\right]^{-1}\left[\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{L}^{t}}\right]\Bigg|_{\boldsymbol{L}^t=\boldsymbol{L}^{t+1}=\hat{\boldsymbol{L}}}.$$

Lemma 3. *Let $(\Gamma_1, \Gamma_2) \in \mathbb{C}^2$ denote the eigenvalues of $X(\hat{\boldsymbol{L}})$. If $|\Gamma_i| < 1$ for both i, then $\hat{\boldsymbol{L}}$ is Lyapunov-stable. If $|\Gamma_i| > 1$ for one $i = 1, 2$, then $\hat{\boldsymbol{L}}$ is not Lyapunov-stable.*

Proof. Note that on $\mathbb{R}^2$ all norms are equivalent in terms of convergence. Here, we use the Euclidean norm and denote it by $\|\cdot\|$. We first recall that the eigenvalues are equal to the roots of the characteristic (second-order) polynomial $\mathsf{Det}(X(\hat{\boldsymbol{L}}) - \Gamma I)$. As such, either both roots are real, or they are complex conjugates of each other. In the case of complex eigenvalues, there exists a coordinate system in which $X(\hat{\boldsymbol{L}})$ acts as a rotation followed by a multiplication with $|\Gamma_1| = |\Gamma_2|$ ("real canonical form").

We distinguish between two cases to show both statements.

(i) First assume that $\Gamma_1, \Gamma_2 \neq 0$. In this case, the discrete version of the Hartmann-Grobmann Theorem in Zgliczyński et al. (2017) shows that it is sufficient to consider the linearization of the problem, i.e.

$$\boldsymbol{L}^{t+1} - \hat{\boldsymbol{L}} = X(\hat{\boldsymbol{L}})(\boldsymbol{L}^{t} - \hat{\boldsymbol{L}}).$$

If $|\Gamma_1|, |\Gamma_2| < 1$, let $\delta = \varepsilon$ and note that

$$\|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \varepsilon \quad \Longrightarrow \quad \|X(\hat{\boldsymbol{L}})^t\,(\boldsymbol{L}^0 - \hat{\boldsymbol{L}})\| \leq \max\{|\Gamma_1|, |\Gamma_2|\}^t \|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \varepsilon$$

for all $t > 0$. Hence, $\hat{\boldsymbol{L}}$ is Lyapunov-stable.

Next, assume $|\Gamma_1| > 1$. If Γ_1 is real, let $\boldsymbol{v}^1$ be the corresponding unit eigenvector. If it is complex, let $\boldsymbol{v}^1$ be an arbitrary vector of length one. Assume that $\hat{\boldsymbol{L}}$ is Lyaponuv-stable for some ε and δ. Let $\boldsymbol{L}^0 = \hat{\boldsymbol{L}} + \delta' \boldsymbol{v}^1$

for $\delta' < \delta$, so that $||\boldsymbol{L}^0 - \hat{\boldsymbol{L}}|| < \delta$. Note that

$$\|X(\hat{\boldsymbol{L}})^t(\boldsymbol{L}^0 - \hat{\boldsymbol{L}})\| \quad = \|\delta'\Gamma_1^t \boldsymbol{v}^1\| = \delta' \cdot |\Gamma_1|^t \|\boldsymbol{v}^1\| \xrightarrow{t\to\infty} \infty.$$

In other words, there exists a large enough t such that $\|X(\hat{\boldsymbol{L}})^t(\boldsymbol{L}^0 - \hat{\boldsymbol{L}})\| > \varepsilon$, which contradicts the assumption of $\hat{\boldsymbol{L}}$ being Lyapunov-stable.

(ii) Next, assume that $\Gamma_2 = 0$ and denote the corresponding eigenvector $\boldsymbol{v}^2$. Then, Γ_1 is real. Let the corresponding unit eigenvector be $\boldsymbol{v}^1$. Then, we can write $\boldsymbol{L}^t - \hat{\boldsymbol{L}} = a_1^t \boldsymbol{v}^1 + a_2^t \boldsymbol{v}^2$ for any t and write the system in the coordinates $\boldsymbol{v}^1, \boldsymbol{v}^2$ as follows:

$$\boldsymbol{L}^{t+1} - \hat{\boldsymbol{L}} = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & 0 \end{pmatrix} (\boldsymbol{L}^t - \hat{\boldsymbol{L}}) + o\big(\|\boldsymbol{L}^t - \hat{\boldsymbol{L}}\|\big).$$

First, consider the case $|\Gamma_1| < 1$ and fix an $\varepsilon > 0$. Then, for any $\eta \in (|\Gamma_1|, 1)$ and for a sufficiently small $\delta < \varepsilon$, whenever $\|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \delta$, $\|\boldsymbol{L}^{t+1} - \hat{\boldsymbol{L}}\| \leq \eta^t \|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \delta < \varepsilon$ by induction over t. Hence, $\hat{\boldsymbol{L}}$ is Lyapunov stable.

Next, assume $|\Gamma_1| > 1$ and assume that $\hat{\boldsymbol{L}}$ is Lyapunov-stable. First, note that for all $\eta \in (1, |\Gamma_1|)$, $\|\boldsymbol{L}^t - \hat{\boldsymbol{L}}\| < \varepsilon$ with $\varepsilon > 0$ sufficiently small and coordinates with respect to the basis $\{\boldsymbol{v}^1, \boldsymbol{v}^2\}$,

$$\begin{aligned}(L_1^{t+1} - \hat{L}_1)^2 - (L_2^{t+1} + \hat{L}_2)^2 &= \Gamma_1^2 (L_1^t - \hat{L}_1)^2 + o\big(\|\boldsymbol{L}^t - \hat{\boldsymbol{L}}\|^2\big) \\ &\geq \eta^2 \big((L_1^t - \hat{L}_1)^2 - (L_2^t - \hat{L}_2)^2\big).\end{aligned} \tag{18}$$

Consider such an $\varepsilon > 0$, the $\delta > 0$ from the definition of Lyapunov stability, and any $\|\boldsymbol{L}^0 - \hat{\boldsymbol{L}}\| < \min\{\delta, \varepsilon\}$ with $(L_1^0 - \hat{L}_1)^2 - (L_2^0 - \hat{L}_2)^2 > 0$. Then, the estimate (18) holds for all t. Hence, $\lim_{t\to\infty} (L_1^t - \hat{L}_1)^2 - (L_2^t - \hat{L}_2)^2 = \infty$. This contradicts the assumption that $\|\boldsymbol{L}^t - \hat{\boldsymbol{L}}\| < \varepsilon$ for all t. Thus, $\hat{\boldsymbol{L}}$ cannot be Lyapunov stable. □

Next, we apply these findings to the dynamic system from Equation (1) more concretely. To this end we introduce some more notation. At a steady

state with composition ϕ, the group-i labor force is given by the mass of individuals with talent above $\hat{x}_i(\phi)$,

$$\hat{L}_1(\phi) := b(1 - F(\hat{x}_1(\phi))) \quad \text{and} \quad \hat{L}_2(\phi) := (1-b)(1 - F(\hat{x}_2(\phi))). \tag{19}$$

Since the mentorship boost increases with representation, $\hat{L}_1'(\phi) \geq 0 \geq \hat{L}_2'(\phi)$ for all compositions $\phi \in (0,1)$. We denote total size by $\hat{L}(\phi) := \hat{L}_1(\phi) + \hat{L}_2(\phi)$. With a slight abuse of notation we use $\hat{L}_i$ to denote both the functions (19) which is defined for arbitrary compositions and the steady state labor force participation. We recover the following stability conditions.

Lemma 4. *A mixed steady state of composition $\hat{\phi} \in (0,1)$ and size $\hat{L}(\hat{\phi})$ is stable if both of the following conditions hold:*

$$0 > -\hat{L}(\hat{\phi}) + (1-\hat{\phi})\hat{L}_1'(\hat{\phi}) - \hat{\phi}\hat{L}_2'(\hat{\phi}) \tag{20}$$

$$\begin{aligned} 0 < {} & 2\hat{L}(\hat{\phi})^2 - \hat{L}(\hat{\phi})\big((1-\hat{\phi}+m_3(\hat{\phi}))\hat{L}_1'(\hat{\phi}) - (\hat{\phi}+m_3(1-\hat{\phi})\hat{L}_2'(\hat{\phi})\big) \\ & + \hat{L}_1'(\hat{\phi})\hat{L}_2'(\hat{\phi})\big(\hat{\phi}(m_1(\hat{\phi}) + m_2(1-\hat{\phi})) + (1-\hat{\phi})(m_1(1-\hat{\phi}) + m_2(\hat{\phi}))\big). \end{aligned} \tag{21}$$

If one of the inequalities is reversed, the steady state is unstable.

Proof. At the steady state labor force $\hat{\boldsymbol{L}}^t = \hat{\boldsymbol{L}}^{t+1} = (\hat{\phi}\hat{L}(\hat{\phi}), (1-\hat{\phi})\hat{L}(\hat{\phi}))$, we can write the matrix

$$X(\hat{\phi}\hat{L}(\hat{\phi}), (1-\hat{\phi})\hat{L}(\hat{\phi})) := -\left[\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{L}^{t+1}}\right]^{-1}\left[\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{L}^{t}}\right]\Big|_{\boldsymbol{L}^t = \boldsymbol{L}^{t+1} = (\hat{\phi}\hat{L}(\hat{\phi}), (1-\hat{\phi})\hat{L}(\hat{\phi}))}$$

as

$$-\begin{bmatrix} 1 - \hat{L}_1'(\hat{\phi})\frac{m_3(\hat{\phi})}{\hat{L}(\hat{\phi})} & -\hat{L}_1'(\hat{\phi})\frac{m_4(\hat{\phi})}{\hat{L}(\hat{\phi})} \\ \hat{L}_2'(\hat{\phi})\frac{m_4(1-\hat{\phi})}{\hat{L}(\hat{\phi})} & 1 + \hat{L}_2'(\hat{\phi})\frac{m_3(1-\hat{\phi})}{\hat{L}(\hat{\phi})} \end{bmatrix}^{-1}\begin{bmatrix} -\hat{L}_1'(\hat{\phi})\frac{m_1(\hat{\phi})}{\hat{L}(\hat{\phi})} & -\hat{L}_1'(\hat{\phi})\frac{m_2(\hat{\phi})}{\hat{L}(\hat{\phi})} \\ \hat{L}_2'(\hat{\phi})\frac{m_2(1-\hat{\phi})}{\hat{L}(\hat{\phi})} & \hat{L}_2'(\hat{\phi})\frac{m_1(1-\hat{\phi})}{\hat{L}(\hat{\phi})} \end{bmatrix}.$$

Recall that the possibly complex eigenvalues Γ_1, Γ_2 are the roots of the characteristic polynomial

$$\mathsf{Det}(X(\hat{\phi}\hat{L}(\hat{\phi}), (1-\hat{\phi})\hat{L}(\hat{\phi})) - \Gamma I) = \frac{1}{\gamma_2}(\gamma_0 + \gamma_1\Gamma + \gamma_2\Gamma^2)$$

with

$$\begin{aligned}
\gamma_0 &= -\hat{L}_1'(\hat{\phi})\hat{L}_2'(\hat{\phi})\big(m_1(\hat{\phi})m_1(1-\hat{\phi}) - m_2(\hat{\phi})m_2(1-\hat{\phi})\big) \geq 0 \\
\gamma_1 &= -\hat{L}(\hat{\phi})\ (\hat{L}_1'(\hat{\phi})m_1(\hat{\phi}) - \hat{L}_2'(\hat{\phi})m_1(1-\hat{\phi})) \\
&\quad - \hat{L}_1'(\hat{\phi})\hat{L}_2'(\hat{\phi})\big(m_1(\hat{\phi})m_3(1-\hat{\phi}) + m_1(1-\hat{\phi})m_3(\hat{\phi}) \\
&\qquad\qquad - m_2(\hat{\phi})m_4(1-\hat{\phi}) - m_2(1-\hat{\phi})m_4(\hat{\phi})\big) \leq 0 \\
\gamma_2 &= \hat{L}(\hat{\phi})^2 - L\big(\hat{L}_1'(\hat{\phi})m_3(\hat{\phi}) - \hat{L}_2'(\hat{\phi})m_3(1-\hat{\phi})\big) \\
&\quad - \hat{L}_1'(\hat{\phi})\hat{L}_2'(\hat{\phi})(m_3(\hat{\phi})m_3(1-\hat{\phi}) - m_4(\hat{\phi})m_4(1-\hat{\phi})\big) > 0
\end{aligned}$$

Since $\hat{L}_1'(\hat{\phi}) \geq 0 \geq \hat{L}_2'(\hat{\phi})$ and $\hat{L}(\hat{\phi}) = \hat{L}_1(\hat{\phi}) + \hat{L}_2(\hat{\phi}) > 0$ in any steady state, the signs on the parameters all follow by Lemma 1. Consequently, this is an upward sloping parabola that is nonnegative and nonincreasing at $\Gamma = 0$. If the function value at $\Gamma = 1$ is nonpositive, there exists a root (and hence an eigenvalue) weakly greater than 1. Thus, a necessary condition for stability is that

$$\gamma_0 + \gamma_1 + \gamma_2 = \hat{L}(\hat{\phi})^2 - \hat{L}(\hat{\phi})\ ((1-\phi)\hat{L}_1'(\phi) - \phi\hat{L}_2'(\phi)) > 0,$$

which after multiplication with $1/\hat{L}(\hat{\phi}) > 0$ yields Equation (20). Further, if the characteristic polynomial is nonincreasing and positive at $\Gamma = 1$, the vertex of the parabola and any real roots are at least one 1. (The vertex corresponds to the real part of any complex eigenvalues.) Thus, another necessary condition is that the derivative at $\Gamma = 1$ is positive, $\gamma_1 + 2\gamma_2 > 0$, as in Equation (21).

If both conditions (20) and (21) hold, the vertex of the polynomial and any real eigenvalues are strictly contained between 0 and 1. Any complex eigenvalues $\Gamma = a \pm bi$ with $b > 0$, solve

$$\begin{cases} \gamma_0 + a\gamma_1 + (a^2 - b^2)\gamma_2 = 0 \\ b\gamma_1 + 2ab\gamma_2 = 0 \qquad \Leftrightarrow \gamma_1 = -2\gamma_2 a \end{cases}$$

because the real and imaginary part of the quadratic function at those values must be zero where $|\Gamma_i| = \sqrt{a^2 + b^2}$. Hence, $|\Gamma_i| < 1$ if and only if $a^2 + b^2 =$

$\frac{\gamma_0}{\gamma_2} < 1$ which is equivalent to $\gamma_2 - \gamma_0 > 0$. This follows from $\gamma_0 + \gamma_1 + \gamma_2 > 0$ and $\gamma_1 + 2\gamma_2 > 0$, so it follows from Equation (20) and Equation (21). Hence, these two conditions are necessary and sufficient for $|\Gamma_i| < 1$.

Taken together, we have shown that Equations (20) and (21) are both necessary and jointly sufficient to ensure that $|\Gamma_i| < 1$ for both $i = 1, 2$; as a result, the steady state is stable. It also follows that if either of the inequalities Equation (20) or Equation (21) is strictly reversed, that there is an eigenvalue with $|\Gamma_i| > 1$, which implies that the steady state is not stable. □

For the proof of Proposition 1, it is useful to define the *majority over-supply* as

$$\Psi(\phi) := (1 - \phi)\hat{L}_1(\phi) - \phi\hat{L}_2(\phi), \tag{22}$$

noting that (17) is satisfied at $(L_1, L_2) = (\phi L, (1-\phi)L)$ if and only if $\Psi(\phi) = 0$.

Proof of Proposition 1. We prove each claim in turn.

(a) Let $\underline{\phi} := \min\{\phi \in [0, 1] \mid \mu(\phi) \geq c - \pi - \bar{x}_F\}$ denote the minimal own-group share to ensure participation of most able individuals. In a homogeneous labor force $\phi \in \{0, 1\}$, the dominant group is always willing to invest because $\underline{\phi} < 0.5$ by Property (F1). A homogeneous workforce constitutes a steady state whenever it is a best response for the dominated group not to invest whenever $\underline{\phi} \geq 0$, which is equivalent to Property (hSS). The steady state is stable whenever $\underline{\phi} > 0$, since small enough perturbations maintain the share of the underrepresented group below the threshold.

(b) Recall that we know from Section A.2.1 that the roots of the majority over-supply function Ψ identify the steady states of the economy. The function Ψ is continuous and

$$\Psi(b) = (1 - b)b\left[F(c - 1 - \mu(1 - b)) - F\left(c - 1 - \mu(b)\right)\right] \geq 0 \tag{23}$$

by monotonicity of μ. When Property (hSS) is satisfied, $\hat{L}_1(\underline{\phi}) = 0$ and the change in sign $\Psi(\underline{\phi}) < 0 \leq \Psi(b)$ implies that Ψ admits a root over

$(\bar{\phi}, b]$. When Property (hSS) is not satisfied, then $\Psi(1) = -\hat{L}_2(1) < 0$, and the change in sign between $\Psi(b) \geq 0 > \Psi(1)$ ensures a root over $[b, 1)$.

(c) Assume that the senior workforce is dominated by the majority, $L_1 > L_2$. We first assume by contradiction that the opposite is true for the junior workforce, $\ell_1 \leq \ell_2$. In that case, majority juniors receive better mentoring than minority juniors,

$$\tilde{\mu}(L_1, L_2, \ell_1, \ell_2) \overset{\text{(M4)}}{\geq} \tilde{\mu}(L_1, L_2, \ell_2, \ell_1) \overset{\text{(M3)}}{\geq} \tilde{\mu}(L_2, L_1, \ell_2, \ell_1).$$

Since $b > 1 - b$, Equation (1) then implies that the individually rational participation is larger for the majority than for the minority, contradicting the assumption that $\ell_1 \leq \ell_2$. As a result, the junior workforce is also dominated by the majority.

To show that the system eventually settles on a workforce that is not just *dominated by* the majority, but *biased* in favor of the majority, we now show that there are no steady states of composition $(0.5, b]$. To do so, note that for $\phi > 0.5$, we have $\hat{x}_1(\phi) < \hat{x}_2(\phi)$. Hence, for $\phi \in (0.5, b]$ we have

$$\Psi(\phi) = (1-\phi)b(1-F(\hat{x}_1(\phi))-\phi(1-b)(1-F(\hat{x}_2(\phi)) > (b-\phi)(1-F(\hat{x}_2(\phi)) > 0.$$

Since Ψ admits no root over that range, the system converges to a steady state that is biased in favor of the majority.

(d) First, we show that for any $\delta > 0$, there exists $Q > 0$ such that the economy admits a stable steady state with composition $\phi \in [b, b + \delta)$ whenever $q > Q$. Indeed, let $\bar{L} = 1 - F(c - \pi - 1)$ denote the size of the labor force if everyone receives the maximal mentoring boost. Since $\lim_{q\to\infty} \mu_q(b + \delta) = 1$ by Property (M5), note that

$$\lim_{q\to\infty} \hat{L}_1(b + \delta) = b\bar{L} \quad \text{and} \quad \lim_{q\to\infty} \hat{L}_2(b + \delta) = (1 - b)\bar{L} \tag{24}$$

and hence

$$\lim_{q\to\infty} \Psi(b+\delta) = (1-b-\delta)b\bar{L} - (b+\delta)(1-b)\bar{L} = -\delta\bar{L} < 0.$$

Since Ψ is continuous and $\Psi(b) \geq 0$ by Equation (23), this implies a downward crossing over $[b, b+\delta)$ for q larger than some threshold Q_1 (i.e., $\Psi(\phi-\varepsilon) > \Psi(\phi) = 0 > \Psi(\phi+\varepsilon)$ for a $\phi \in [b, b+\delta)$ and all $\varepsilon > 0$ sufficiently small). In addition, the downward slope $\Psi'(\phi) < 0$ is equivalent to the first stability condition in Lemma 4. In order to show the second stability condition (21), note that the convergence of $\mu_q'(\phi) \to 0$ is uniform over $[b, b+\delta]$ by Dini's Theorem[17] and in turn implies uniform convergence of $\hat{L}_i'(\phi) \to 0$ for either group i. By Lemma 1, $|m_k(\phi)|$ admit an upper bound M that is independent of k, ϕ and q. As a consequence, the right side of Equation (21) converges to $2\bar{L}^2 > 0$, implying that there exists Q_2 such that the economy admits a stable steady state within $[b, b+\delta)$ whenever $q > \max\{Q_1, Q_2\}$.

Finally, we rule out any other steady states for q large enough. By Property (M5), $\lim_{q\to\infty} \hat{L}_1(\delta) = b\bar{L}$ and $\lim_{q\to\infty} \hat{L}_2(\delta) = (1-b)\bar{L}$. By the definition of $\bar{L}$, both sequences approach the limit from below. Hence, there exists a Q_3 so that for all $q > Q_3$,

$$\hat{L}_1(\delta) \in \left(\left(b - \frac{\delta}{b+\delta}\right)\bar{L}, b\bar{L}\right] \quad \text{and} \quad \hat{L}_2(1-\delta) \in \left(\left(1-b-\frac{\delta}{b+\delta}\right)\bar{L}, (1-b)\bar{L}\right].$$

Since μ is increasing, we have $\hat{L}_1(\phi) \in [\hat{L}_1(\delta), b\bar{L}]$ and $\hat{L}_2(\phi) \in [\hat{L}_2(1-\delta), (1-b)\bar{L}]$ for all $\phi \in [\delta, 1-\delta]$. In turn, these bounds imply

$$\Psi(\phi) \in \left((b-\phi)\bar{L} - (1-\phi)\frac{\delta}{b+\delta}\bar{L}, (b-\phi)\bar{L} + \phi\frac{\delta}{b+\delta}\bar{L}\right).$$

At any root of Ψ, this range must contain zero. Rewriting that condition

[17] Dini's Theorem states that if a monotone sequence of continuous functions converges pointwise on a compact space, and if the limit function is also continuous, then the convergence is uniform.

in terms of ϕ, we obtain the equivalent expression for any root ϕ:

$$\phi \in \left(b - \delta + \delta\frac{2b-1}{b}, b + \delta\right) \subseteq (b - \delta, b + \delta).$$

In other words, for $q > Q_3$, the only roots of Ψ are either almost homogeneous $\phi \in [0, \delta) \cup (1 - \delta, 1]$ or almost fair $\phi \in (b - \delta, b + \delta)$.

(e) First, we show that there exists a $\Lambda_1 \geq 0$ so that for all $\lambda > \Lambda_1$, there is a steady state in $[b, b + \delta)$. To this end, note that Lemma 2 implies that

$$\lim_{\lambda\to\infty} \frac{\hat{L}_1(\phi)}{\hat{L}_2(\phi)} = \lim_{\lambda\to\infty} \frac{b(1 - F(\hat{x}_1(\phi)))}{(1-b)(1 - F(\hat{x}_2(\phi)))} = \frac{b}{1-b} \qquad \forall \phi \in [0, 1] \quad (25)$$

and hence

$$\lim_{\lambda\to\infty} \frac{\Psi(b+\delta)}{\hat{L}_1(b+\delta) + \hat{L}_2(b+\delta)} = \lim_{\lambda\to\infty} \frac{(1-b-\delta)\frac{\hat{L}_1(b+\delta)}{\hat{L}_2(b+\delta)} - (b+\delta)}{\frac{\hat{L}_1(b+\delta)}{\hat{L}_2(b+\delta)} + 1} = -\delta < 0.$$

In other words, there exists Λ_1 such that this ratio is negative for all $\lambda > \Lambda_1$. This in turn implies $\Psi(b + \delta) < 0 \overset{(23)}{\leq} \Psi(b)$, and thus Ψ has a downward crossing root over $[b, b + \delta)$ for all $\lambda > \Lambda_1$, i.e. there is a steady state which satisfies the first condition (20) of Lemma 4.

Next, we show that there exists a $\Lambda_2 \geq 0$ so that for $\lambda > \max\{\Lambda_1, \Lambda_2\}$, all such steady states in $[b, b + \delta)$ are stable. Since it is a downward crossing, we only need to show that the second condition Equation (21) of Lemma 4 is satisfied. The continuous functions $\mu'(\phi)$ and $m_k(\phi)$ are all bounded over the compact interval $[b, b + \delta]$. Property (F2) then implies that

$$\lim_{\lambda\to\infty} \frac{\hat{L}_1'(\phi)}{\hat{L}_1(\phi)} = \lim_{\lambda\to\infty} \frac{F'(\hat{x}_1(\phi))\mu'(\phi)}{1 - F(\hat{x}_1(\phi))} \overset{(F2)}{=} 0$$

for all $\phi \in [b, b + \delta]$, and the convergence is uniform by Dini's Theorem. Dividing Equation (21) by $L^2 > 0$, note that the right side converges to 2 as $\lambda \to \infty$. In other words, there exists Λ_2 big enough such that the economy

admits a stable steady state within $[b, b+\delta)$ whenever $\lambda > \max\{\Lambda_1, \Lambda_2\}$.

Furthermore, the convergence in Equation (25) is uniform by Dini's Theorem, implying that there exists Λ_3 large enough such that

$$\frac{\Psi(\phi)}{\hat{L}_2(\phi)} \in \left((1-\phi)\left(\frac{b}{1-b} - \varepsilon\right) - \phi, (1-\phi)\left(\frac{b}{1-b} + \varepsilon\right) - \phi\right) \qquad \forall \phi \in [0,1]$$

At any root of Ψ, this range must contain zero. Letting $\varepsilon = \frac{\delta}{(1-b)(1-b+\delta)}$ and rewriting the condition in terms of ϕ, we obtain the equivalent expression

$$\phi \in \left(b - \delta, b + \frac{1-b}{1-b+2\delta}\delta\right) \subseteq (b-\delta, b+\delta).$$

In other words, for $\lambda > \Lambda_3$, the only roots of Ψ are almost fair with composition $\phi \in (b-\delta, b+\delta)$. □

A.2.2 Optimal labor force composition

Let $L^* : [0,1] \rightrightarrows [0,\infty)$ and $S^* : [0,1] \to \mathbb{R}$ be defined from Equation (3) as

$$L^*(\phi) = \arg\max_{L \geq 0} S(\phi, L) \qquad \text{and} \qquad S^*(\phi) = \max_{L \geq 0} S(\phi, L).$$

We refer to the cutoffs under composition ϕ and labor force size L as $\hat{x}_1 = F^{-1}\big(1 - \frac{\phi}{b}L\big)$ and $\hat{x}_2 = F^{-1}\big(1 - \frac{1-\phi}{1-b}L\big)$. When $L = L^*(\phi)$ is optimal for composition ϕ, we write the cutoffs as x_1^* and x_2^*. The first result in Lemma 5 establishes uniqueness of the maximizer, and we abuse notation by referring to its unique element as $L^*(\phi)$.

Lemma 5. *The functions L^* and S^* satisfy the following properties:*

(a) $L^*(\phi)$ is singleton-valued and strictly positive for all $\phi \in [0,1]$.

(b) At the optimal composition $\phi^* = \arg\max_\phi S^*(\phi)$, a positive mass of each group abstains from participation, $x_1^*, x_2^* > \underline{x}_F$.

(c) Whenever $x_1^*, x_2^* > \underline{x}_F$, optimal surplus S^* and its derivative $S^{*\prime}$ can be written as

$$S^*(\phi) = b\int_{x_1^*}^{\bar{x}}(1-F(x))dx + (1-b)\int_{x_2^*}^{\bar{x}}(1-F(x))dx, \tag{26}$$

$$S^{*\prime}(\phi) = L^*(\phi)\left(\mu(\phi) - \mu(1-\phi) + \phi\mu'(\phi) - (1-\phi)\mu'(1-\phi) - x_1^* - x_2^*\right) \tag{27}$$

where x_1^* and x_2^* solve

$$0 = \pi - c + \phi\mu(\phi) + (1-\phi)\mu(1-\phi) + \phi x_1^* + (1-\phi)x_2^* \tag{28}$$

$$L^*(\phi) = \frac{b}{\phi}(1-F(x_1^*)) = \frac{1-b}{1-\phi}(1-F(x_2^*)). \tag{29}$$

(d) S^* has the following properties:

$$S^*(1) = b\int_{c-\pi-\mu(1)}^{\bar{x}}(1-F(x))dx \tag{30}$$

$$S^*(b) = \int_{c-\pi-b\mu(b)-(1-b)\mu(1-b)}^{\bar{x}}(1-F(x))dx \tag{31}$$

$$S^{*\prime}(1) = b(1-F(c-\pi-\mu(1))\left(c-\pi-\mu(0)+\mu'(1)-\bar{x}_F\right) \tag{32}$$

$$S^{*\prime}(b) = L^*(b)(\mu(b)+b\mu'(b)-\mu(1-b)-(1-b)\mu'(1-b)) \tag{33}$$

$$S^{*\prime}(0.5) = L^*(0.5)(x_1^*(0.5)-x_2^*(0.5)) \geq 0, \tag{34}$$

where $x_1^*(\phi) = F^{-1}(1-\frac{\phi}{b}L^*(\phi))$ and $x_2^*(\phi) = F^{-1}(1-\frac{1-\phi}{1-b}L^*(\phi))$.

At any steady state $(\hat{\phi}, \hat{L})$,

$$S(\hat{\phi},\hat{L}) = b\int_{c-\pi-\mu(\hat{\phi})}^{\bar{x}}(1-F(x))dx + (1-b)\int_{c-\pi-\mu(1-\hat{\phi})}^{\bar{x}}(1-F(x))dx \tag{35}$$

$$S^{*\prime}(\hat{\phi}) = \hat{L}\left(\hat{\phi}\mu'(\hat{\phi}) - (1-\hat{\phi})\mu'(1-\hat{\phi})\right). \tag{36}$$

Proof. We prove each claim in turn:

(a) First, note that we can write

$$\begin{aligned} S(\phi, L) &= b \int_{1-\frac{\phi}{b}L}^{1} \pi - c + F^{-1}(y) + \mu(\phi) dy \\ &\quad + (1-b) \int_{1-\frac{1-\phi}{1-b}L}^{1} \pi - c + F^{-1}(y) + \mu(1-\phi) dy. \end{aligned}$$

Hence, we have

$$\frac{\partial}{\partial L} S(\phi, L) = \pi - c + M(\phi) + \phi \hat{x}_1 + (1-\phi)\hat{x}_2 \tag{37}$$

$$\frac{\partial}{\partial \phi} S(\phi, L) = L\left(M'(\phi) + \hat{x}_1 - \hat{x}_2\right) \tag{38}$$

where $M(\phi) = \phi\mu(\phi) + (1-\phi)\mu(1-\phi)$ is as defined in (4), but we omit the subscript in this proof. Furthermore, surplus is strictly concave in L for any ϕ,

$$\frac{\partial^2}{\partial L^2} S(\phi, L) = -\frac{\phi^2}{b} \frac{1}{F'(F^{-1}(1-\frac{\phi}{b}L))} - \frac{(1-\phi)^2}{1-b} \frac{1}{F'(F^{-1}(1-\frac{1-\phi}{1-b}L))} < 0,$$

ensuring a unique solution.

(b) Positive size L is always optimal since

$$\frac{\partial}{\partial L} S(\phi, 0) = \pi - c + \bar{x}_F + M(\phi) \geq \pi - c + \bar{x}_F + \mu(0.5) > 0$$

by Property (F1). When a compositions severely over-represents one group i and talent dispersion is high, it may be optimal to set $x_i^* = \underline{x}_F$, as this allows for the participation of opposite-group members with high individual surplus. Note, however, that at the cutoffs (x_1^*, x_2^*) that are optimal

under composition ϕ, Equations (37) and (38) jointly imply

$$\begin{aligned}\frac{\partial}{\partial\phi}S\left(\phi,L\right)\big|_{L=L^*(\phi)} &= L\left(M'(\phi)+x_1^*-x_2^*\right)\\ &= \frac{L}{1-\phi}\Bigg(\underbrace{(1-\phi)M'(\phi)+M(\phi)+x_1^*+\pi-c}_{(T_1)}-\frac{\partial}{\partial L}S\left(\phi,L\right)\Bigg) \quad (39)\\ &= -\frac{L}{\phi}\Bigg(\underbrace{\phi M'(1-\phi)+M(1-\phi)+x_2^*+\pi-c}_{(T_2)}-\frac{\partial}{\partial L}S\left(\phi,L\right)\Bigg), \quad (40)\end{aligned}$$

where we used the fact that $M(\phi)=M(1-\phi)$. If the optimal labor size $L^*(\phi)$ employs all members of group i, then $\frac{\partial}{\partial L}S\left(\phi,L\right)\big|_{L=L^*(\phi)}\geq 0$ and $x_i^*=\underline{x}_F$. By assumption Property (F1) on the lower bound $\underline{x}_F$, the term T_i in the expression above is strictly negative. When $i=1$, Equation (39) implies that $\frac{\partial}{\partial\phi}S\left(\phi,L\right)\big|_{L=L^*(\phi)}<0$, and when $i=2$, Equation (40) implies $\frac{\partial}{\partial\phi}S\left(\phi,L\right)\big|_{L=L^*(\phi)}>0$. In either case, a marginal adjustment in composition strictly improves surplus. This ensures that all cutoffs are interior at the optimal composition.

(c) For any interior maximum, we have $0=\frac{\partial}{\partial L}S(\phi,L)\big|_{L=L^*(\phi)}$, which imposes condition (28) on the optimal cutoffs x_1^* and x_2^*. The second condition (29) merely restates the definition in Equation (3).

(d) By Fubini's Theorem, we can write

$$\int_{x_i^*}^{\bar{x}}(x-x_i^*)F'(x)dx=\int_{x_i^*}^{\bar{x}}\int_{x_1^*}^{x}F'(x)dtdx=\int_{x_i^*}^{\bar{x}}\int_{t}^{\bar{x}}F'(x)dxdt=\int_{x_i^*}^{\bar{x}}(1-F(x))dx$$

and hence

$$\begin{aligned}
S(\phi, L) &= L\left(\pi - c + \phi\mu(\phi) + (1-\phi)\mu(1-\phi)\right) \\
&\quad + bx_1^*(1 - F(x_1^*)) + (1-b)x_2^*(1 - F(x_2^*)) \\
&\quad + b\int_{x_1^*}^{\bar{x}} (x - x_1^*)F'(x)dx + (1-b)\int_{x_2^*}^{\bar{x}} (x - x_2^*)F'(x)dx \\
&= L\left(\pi - c + \phi\mu(\phi) + (1-\phi)\mu(1-\phi) + \phi x_1^* + (1-\phi)x_2^*\right) \\
&\quad + b\int_{x_1^*}^{\bar{x}} (1 - F(x))dx + (1-b)\int_{x_2^*}^{\bar{x}} (1 - F(x))dx.
\end{aligned}$$

This implies Equation (35) and, together with (28), we obtain Equation (26) By the Envelope Theorem, we obtain $S^{*\prime}(\phi) = \frac{\partial S}{\partial \phi}(\phi, L^*(\phi))$ and hence (27).

At composition $\phi = 1$, we have $x_2^* = \bar{x}_F$. By Equation (28), $x_1^* = c - \pi - \mu(1)$. By Equation (29), $L^*(1) = b(1 - F(c - \pi - \mu(1))$. Plugging this into Equation (26) yields (30), and plugging it into Equation (27) yields (32).

At composition $\phi = b$, we have $x_1^* = x_2^* = F^{-1}(1 - L^*(b)) \overset{(28)}{=} c - \pi - \phi\mu(b) - (1-b)\mu(1-b)$. Plugging this into Equation (26) yields (31), and plugging it into Equation (27) yields (33).

At any steady state $(\hat{\phi}, \hat{L})$, Equation (28) holds since the participation constraint (IR) is binding at x_1^* and x_2^*, and thus $\hat{L} = L^*(\hat{\phi})$ and $x_1^* - x_2^* = \mu(1-\hat{\phi}) - \mu(\hat{\phi})$. Plugging this into Equation (26) yields (35), and plugging it into Equation (27) yields (36).

□

Proof of Proposition 2. In Proposition 1, we establish the existence of a stable steady state arbitrarily close to composition b for large enough q or λ. The surplus of that steady state is given by Equation (35) in Lemma 5 and converges to

$$\lim_{q\to\infty} b\int_{c-\pi-\mu(b)}^{\bar{x}} (1 - F(x))dx + (1-b)\int_{c-\pi-\mu(1-b)}^{\bar{x}} (1 - F(x))dx = \int_{c-\pi-1}^{\bar{x}} (1 - F(x))dx > 0$$

as $q \to \infty$. Conversely, the surplus of the best homogeneous steady state $\hat{\phi} = 1$ converges to

$$\lim_{q\to\infty} b \int_{c-\pi-\mu(1)}^{\bar{x}} (1 - F(x))dx = b \int_{c-\pi-1}^{\bar{x}} (1 - F(x))dx$$

by Equation (31). Since $b < 1$, the nearly fair mixed stable steady state eventually yields a higher surplus than any homogeneous steady state.

As for $\lambda \to \infty$, the results follows simply because there are no homogeneous steady states for λ large enough by Proposition 1(e). □

Proof of Proposition 3. We prove each claim in turn:

(a) Property (F1) implies that $\mu_q(0.5) > c - \pi - \bar{x}_F$ for at least some $q > 0$. Let us denote such a q by $\tilde{q}$. Since $\lim_{q\to 0} \mu_q(0.5) = 0$ by Property (M5) and $0 < c - \pi - \bar{x}_F$, the Intermediate Value Theorem guarantees the existence of $q_{0.5} \in (0, \tilde{q})$ such that $\mu_{q_{0.5}}(0.5) = c - \pi - \bar{x}_F$. Consider now what happens when $q < q_{0.5}$. Since μ is strictly increasing in ϕ and q, any participating member of the minority group has negative individual surplus, $0 > \pi - c + x + \mu_q(1 - \phi)$. Excluding group 2 from the workforce also improves the mentorship boost for the majority, and hence the optimal labor force is homogeneous. Since $\mu_{q_{0.5}}(1) > c - \pi - \bar{x}_F$, there exists a nonempty range of $q < q_{0.5}$ where the most talented majority members generate positive surplus under $\phi = 1$, and thus ensure a strictly positive size of the workforce.

(b) First, we establish that for $b > 0.5$, the optimal workforce is always weakly dominated by the majority, $\phi^* \geq 0.5$. Indeed, using a change of variables, we can write the surplus from Equation (3) as

$$\begin{aligned} S(\phi, L) = L(\pi - c + x + \phi\mu(\phi) + (1 - \phi)\mu(1 - \phi)) \\ + L \int_0^{\phi} F^{-1}\left(1 - \frac{y}{b}L\right) dy + L \int_0^{1-\phi} F^{-1}\left(1 - \frac{y}{1-b}L\right) dy. \end{aligned}$$

Consider now any labor force $(\phi L, (1 - \phi)L)$ for some $\phi < 0.5$, and com-

pare its surplus to that of labor force $((1-\phi)L, \phi L)$. The two quantities differ only in the range of integration for the last two terms. Since $F^{-1}\left(1-\frac{q}{b}L\right) > F^{-1}\left(1-\frac{q}{1-b}L\right)$ pointwise, surplus is strictly higher for the labor force that is dominated by the majority.

Next, we show that the optimal labor force is biased in favor of the minority for q large enough. We show this in two steps.

First, by the expressions for $S^*(b)$ and $S^*(1)$ in Lemma 5,

$$\lim_{q\to\infty} \frac{S^*(1)}{S^*(b)} = \frac{b\int_{c-\pi-1}^{\bar{x}}(1-F(x))dx}{\int_{c-\pi-1}^{\bar{x}}(1-F(x))dx} = b < 1.$$

Consequently, there exists $Q_1 \in \mathbb{R}$ such that the optimal workforce is mixed for all $q > Q_1$.

Next, we show that there exists a $Q_2 > 0$ so that for $q > \max\{Q_1, Q_2\}$, a fair labor force generates higher surplus than one with a bias in favor of the majority, $S^*(b) > S^*(\phi)$ for all $\phi > b$. To that end, we define

$$\varepsilon := \int_{c-\pi-1}^{\bar{x}}(1-F(x))dx + \int_{x_1^{*\infty}}^{c-\pi-1}(1-F(x_1^{*\infty}))dx - \int_{x_1^{*\infty}}^{\bar{x}}(1-F(x))dx,$$

where $(x_1^{*\infty}, x_2^{*\infty})$ denotes the limiting cutoffs for $q \to \infty$. Note that in any mixed workforce of composition ϕ, $x_1^{*\infty}$ and $x_2^{*\infty}$ satisfy

$$0 = \pi - c + 1 + \phi x_1^{*\infty} + (1-\phi)x_2^{*\infty}, \tag{41}$$

$$\frac{b}{\phi}(1-F(x_1^{*\infty})) = \frac{1-b}{1-\phi}(1-F(x_2^{*\infty})), \tag{42}$$

due to constraints (28) and (29). Whenever $\phi > b$, the limiting cutoff is lower for the majority group, $x_1^{*\infty} < x_2^{*\infty}$, by Equation (42) and by Equation (41), it also follows that $x_1^{*\infty} < c - \pi - 1 < x_2^{*\infty}$. Together with monotonicity of F, this implies that ε is positive,

$$\varepsilon > \int_{c-\pi-1}^{\bar{x}}(1-F(x))dx + \int_{x_1^{*\infty}}^{c-\pi-1}(1-F(x))dx - \int_{x_1^{*\infty}}^{\bar{x}}(1-F(x))dx > 0.$$

The same is true for $\delta := b\frac{\varepsilon}{2}$.

Moreover, since F is bounded and $\mu(\phi) \to 1$ pointwise, there exists $Q_2 \in \mathbb{R}$ such that for all $q > Q_2$

$$\max\left\{\left|\int_{x_i^*}^{x_i^{*\infty}}(1-F(x))dx\right|, \left|\int_{c-\pi-b\mu(b)-(1-b)\mu(1-b)}^{c-\pi-1}(1-F(x))dx\right|\right\}_{i=1,2} < \delta. \tag{43}$$

We will use this to generate an upper bound for the surplus expression (26) for any $\phi > b$. Indeed, note that

$$\int_{x_1^*}^{\bar{x}}(1-F(x))dx < \delta + \int_{c-\pi-1}^{\bar{x}}(1-F(x))dx + \int_{x_1^{*\infty}}^{c-\pi-1}(1-F(x_1^{*\infty}))dx - \varepsilon \tag{44}$$

and similarly,

$$\begin{aligned}\int_{x_2^*}^{\bar{x}}(1-F(x))dx &< \delta + \int_{c-1-\pi}^{\bar{x}}(1-F(x))dx - \int_{c-1-\pi}^{x_2^{*\infty}}\underbrace{(1-F(x))}_{>1-F(x_2^{*\infty})}dx \\ &< \delta + \int_{c-1-\pi}^{\bar{x}}(1-F(x))dx - \int_{c-1-\pi}^{x_2^{*\infty}}(1-F(x_2^{*\infty}))dx. \end{aligned}\tag{45}$$

Note also that by Equations (41) and (42),

$$\begin{aligned} b\int_{x_1^{*\infty}}^{c-\pi-1}(1-F(x_1^{*\infty}))dx \;&=\; b(1-F(x_1^{*\infty}))\overbrace{(c-\pi-1-x_1^{*\infty})}^{\overset{(41)}{=}(1-\phi)(x_2^{*\infty}-x_1^{*\infty})} \\ \overset{(42)}{=} (1-b)(1-F(x_2^{*\infty}))\underbrace{\phi(x_2^{*\infty}-x_1^{*\infty})}_{\overset{(41)}{=}x_2^{*\infty}-(c-\pi-1)} &= (1-b)\int_{c-\pi-1}^{x_2^{*\infty}}(1-F(x_2^{*\infty}))dx. \end{aligned}$$

We conclude that for all $q > Q_2$ and $\phi > b$,

$$\begin{aligned} S^*(\phi) &= b\int_{x_1^*}^{\bar{x}}(1-F(x))dx + (1-b)\int_{x_2^*}^{\bar{x}}(1-F(x))dx && \text{by (26)} \\ &< \delta - b\varepsilon + \int_{c-\pi-1}^{\bar{x}}(1-F(x))dx && \text{by (44) and (45)} \\ &< 2\delta - b\varepsilon + \int_{c-\pi-b\mu(b)-(1-b)\mu(1-b)}^{\bar{x}}(1-F(x))dx && \text{by (43)} \\ &= S^*(b) && \text{by (31).} \end{aligned}$$

This shows that over-representation of the majority is suboptimal for any $q > \max\{Q_1, Q_2\}$. The optimal composition is thus contained in $[0.5, b]$.

For $b > 0.5$, we can rule out the boundaries of this interval. Indeed, Equation (33) implies that $S^{*\prime}(b)$ has the same sign as

$$\mu_q(b) + b\mu_q'(b) - \mu_q(1-b) - (1-b)\mu_q'(1-b).$$

By Property (M6), this is negative for all q above a threshold Q_3. Together with the strict inequality in Equation (34), this implies that the optimal composition is in $(0.5, b)$ whenever $q > \max\{Q_1, Q_2, Q_3\}$.

(c) The optimal composition satisfies the first order condition $S^{*\prime}(\phi)$, which by Equation (27) implies that

$$0 = \mu(\phi) - \mu(1-\phi) + \phi\mu'(\phi) - (1-\phi)\mu'(1-\phi) + x_1^* - x_2^* \tag{46}$$

Since all but the last two terms disappear as $q \to \infty$, the only way both this and the two conditions (41) and (42) can be satisfied is if $x_1^{*\infty} = x_2^{*\infty} = c - \pi - 1$ and $\phi = b$. By continuity of the derivatives, the optimal composition therefore converges to b.

Finally, we consider the impact of high talent dispersion λ on the optimal labor force composition. Our proof relies on three observations: First, a homogeneous workforce is suboptimal for large enough λ. This can most

easily be seen from Equation (32), which shows that $S'(1)$ is negative whenever $x_F^* > c - \pi - \mu(0) + \mu'(1)$. Second, any interior optimum has to solve $S^{*\prime}(\phi) = 0$, which together with Equations (28) and (29) identifies all local extrema. Letting $M(\phi) := \phi\mu(\phi) + (1-\phi)\mu(1-\phi)$, Equations (27) and (28) imply that marginal talent is equal to

$$x_1^* = c - \pi - M(\phi) - (1-\phi)M'(\phi) \qquad \text{and} \qquad x_2^* = c - \pi - M(\phi) + \phi M'(\phi)$$

at the optimal composition ϕ. Both expressions are independent of λ, and in the support of the talent function for large enough λ by Property (F1). Continuity of μ' implies that whenever $M'(b) < 0$, there exists $\delta > 0$ such that $M'(\phi) < 0$ for all $\phi \in [b, b+\delta)$. The optimal composition has to satisfy Equation (29), which yields

$$\phi = \left(1 + \frac{1-b}{b} \frac{1 - F(c - \pi - M(\phi) + \phi M'(\phi))}{1 - F(c - \pi - M(\phi) - (1-\phi)M'(\phi))}\right)^{-1}.$$

When $M'(\phi) < 0$ the right side is strictly smaller than b, ruling out any solutions over $[b, b+\delta)$. Moreover, Equation (15) implies that the right side converges to b as $\lambda \to \infty$. The convergence is uniform by Dini's Theorem. For λ large enough, this rules out any solutions larger than $\phi \geq b + \delta$.

Conversely, whenever $M'(b) > 0$, the same argument rules out solutions over some interval $(b - \delta', b]$ and then restricts crossings to a δ'-ball around b for λ large enough. For high enough talent dispersion, Equation (7) therefore identifies the threshold q that determines the bias of the optimal labor force. □

References

Athey, Susan, Christopher Avery, and Peter Zemsky, "Mentoring and Diversity," *American Economic Review*, 2000, *90* (4), 765–786.

Bayer, Amanda and Cecilia Elena Rouse, "Diversity in the Economics Profession: A New Attack on an Old Problem," *Journal of Economic Perspectives*, November 2016, *30* (4), 221–42.

BBC News, Reality Check Team, "South Africa elections: Who controls the country's business sector?," https://www.bbc.com/news/world-africa-48123937 2019. Accessed: 2020-06-03.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova, "Powerful Women: Does Exposure Reduce Bias?," *The Quarterly Journal of Economics*, 2009, *124* (4), 1497–1540.

Becker, Gary S. and Nigel Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 1979, *87* (6), 1153–1189.

Becker, Gary Stanley, *The economics of discrimination: an economic view of racial discrimination*, University of Chicago, 1957.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, *75* (4), 352–365.

Bertrand, Marianne, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney, "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway," Working Paper 20256, National Bureau of Economic Research June 2014.

Besley, Timothy, Olle Folke, Torsten Persson, and Johanna Rickne, "Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden," *American Economic Review*, August 2017, *107* (8), 2204–42.

Bettinger, Eric P and Bridget Terry Long, "Do faculty serve as role models? The impact of instructor gender on female students," *American Economic Review*, 2005, pp. 152–157.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg, "The Dynamics of Discrimination: Theory and Evidence," July 2018. PIER Working Paper No. 18-016. Available at SSRN: https://ssrn.com/abstract=3235376.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

_ , _ , _ , and _ , "Beliefs about gender," *American Economic Review*, 2019, *109* (3), 739–73.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 2017. Forthcoming.

Card, David and Laura Giuliano, "Universal screening increases the representation of low-income and minority students in gifted education," *Proceedings of the National Academy of Sciences*, 2016, *113* (48), 13678–13683.

Carlana, Michela, "Implicit Stereotypes: Evidence from Teachers Gender Bias*," *The Quarterly Journal of Economics*, 03 2019, *134* (3), 1163–1224.

Carrell, Scott E, Marianne E Page, and James E West, "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Carvalho, Jean-Paul and Bary Pradelski, "Identity and Underrepresentation," December 2018. SSRN working paper. Available at https://ssrn.com/abstract=3299477.

Casas-Arce, Pablo and Albert Saiz, "Women and Power: Unpopular, Unwilling, or Held Back?," *Journal of Political Economy*, 2015, *123* (3), 641–669.

Chung, Kim-Sau, "Role models and arguments for affirmative action," *American Economic Review*, 2000, pp. 640–648.

Coate, Stephen and Glenn C Loury, "Will affirmative-action policies eliminate negative stereotypes?," *The American Economic Review*, 1993, pp. 1220–1240.

Dee, Thomas S., "Teachers, Race, and Student Achievement in a Randomized Experiment," *The Review of Economics and Statistics*, 2004, *86* (1), 195–210.

Dee, Thomas S, "A teacher like me: Does race, ethnicity, or gender matter?," *The American economic review*, 2005, *95* (2), 158–165.

Dee, Thomas S., "Teachers and the Gender Gaps in Student Achievement," *The Journal of Human Resources*, 2007, *42* (3), 528–554.

DeLong, Thomas J, John J Gabarro, and Robert J Lees, "Why mentoring matters in a hypercompetitive world," *Harvard Business Review*, 2008, *86* (1), 115.

Dobbin, Frank and Alexandra Kalev, "Why Diversity Programs Fail," *Harvard Business Review*, 2016, *94* (7-8), 52–60.

Dreher, George F. and Taylor H. Cox Jr., "Race, Gender, and Opportunity: A Study of Compensation Attainment and the Establishment of Mentoring Relationships.," *Journal of Applied Psychology*, 1996, *81* (3), 297 – 308.

Egan, Mary Ellen, "Global Diversity Rankings by Country, Sector and Occupation," in "Diversity & Inclusion: Unlocking Global Potential," New York: Forbes Insight, 2012.

Ellison, Glenn and Ashley Swanson, "The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions," Technical Report, National Bureau of Economic Research 2009.

Fairlie, Robert W, Florian Hoffmann, and Philip Oreopoulos, "A community college instructor like me: Race and ethnicity interactions in the classroom," *The American Economic Review*, 2014, *104* (8), 2567–2591.

Fang, Hanming and Andrea Moro, "Theories of Statistical Discrimination and Affirmative Action: A Survey," in Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds., *Handbook of Social Economics, Vol. 1A*, The Netherlands: North-Holland, 2011, pp. 133–200.

Fryer, Roland G. Jr. and Glenn C. Loury, "Affirmative Action and Its Mythology," *Journal of Economic Perspectives*, September 2005, *19* (3), 147–162.

Herskovic, Bernard and Joao Ramos, "Promoting Educational Opportunities: Long-Run Implications of Affirmative Action in College Admissions," July 2017. SSRN working paper. Available at https://ssrn.com/abstract=2628303.

Ibarra, Herminia, "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm," *Administrative Science Quarterly*, 1992, *37* (3), pp. 422–447.

Jackson, C. Kirabo, "The Effect of Single-Sex Education on Test Scores, School Completion, Arrests, and Teen Motherhood: Evidence from School Transitions," Working Paper 22222, National Bureau of Economic Research May 2016.

Jr, Roland G Fryer, "Belief flipping in a dynamic model of statistical discrimination," *Journal of Public Economics*, 2007, *91* (5-6), 1151–1166.

Kahlenberg, Richard D., Halley Potter, Tanya Katerí Hernández, Haibo Huang, F. Michael Higginbotham, Jennifer Lee, and Angel L. Harris, "Should Affirmative Action Be Based on Income?," *The New York Times*, April 27 2014.

Kahneman, Daniel and Amos Tversky, "Subjective probability: A judgment of representativeness," *Cognitive psychology*, 1972, *3* (3), 430–454.

Kofoed, Michael S et al., "The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at west point," *Journal of Human Resources*, 2019, *54* (2), 430–467.

Leonhardt, David, "Rethinking Affirmative Action," *The New York Times*, October 13 2012.

Matsa, David A. and Amalia R. Miller, "Chipping away at the Glass Ceiling: Gender Spillovers in Corporate Leadership," *The American Economic Review*, 2011, *101* (3), 635–639.

Porter, Catherine and Danila Serra, "Gender differences in the choice of major: The importance of female role models," *American Economic Journal: Applied Economics*, 2019.

Prenovitz, Sarah J., Gary R. Cohen, Ronald G. Ehrenberg, and George H. Jakubson, "An evaluation of the Mellon Mays Undergraduate Fellowship's effect on PhD production at non-UNCF institutions," *Economics of Education Review*, 2016, *53*, 284 – 295.

Restuccia, Diego and Carlos Urrutia, "Intergenerational Persistence of Earnings: The Role of Early and College Education," *American Economic Review*, December 2004, *94* (5), 1354–1378.

Zgliczyński, Piotr et al., "Topological shadowing and the Grobman-Hartman theorem," *Topological Methods in Nonlinear Analysis*, 2017, *50* (2), 757–785.