

# MENTORING AND THE DYNAMICS OF AFFIRMATIVE ACTION\*

Michèle Müller-Itten<sup>†</sup> and Aniko Öry<sup>‡</sup>

November 29, 2018

## Abstract

We study the dynamics of workforce participation when same-group mentoring lowers education costs. Our continuous-time overlapping-generations model considers a majority and a minority population group of identically distributed talent. Under sufficiently decreasing returns to mentoring, and in high-skill sectors, we find that a social planner should enforce an over-representation of minority workers relative to their population share. Such a composition never arises endogenously as a steady state, and thus requires persistent government intervention. We discuss how this intuition qualitatively differs from existing models of workforce composition and the “glass ceiling effect”, and contrast different policy instruments.

**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 and Lisa Kahn for insightful discussions and comments. Financial support from NSF (grant # SES 1227707) is gratefully acknowledged.

<sup>†</sup>University of Notre Dame. Email: michele.muller@nd.edu.

<sup>‡</sup>Yale University. Email: aniko.oery@yale.edu.

# 1 Introduction

This much is known: Mentorship opportunities arise more readily between members of the same race or gender, with important repercussions on professional achievement. As today's graduates turn into tomorrow's mentors, these effects potentially exacerbate with time. But what does this mean for the optimal minority representation in the workforce?

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 surplus. In a nutshell, we assume that labor force evolution is governed by workers' education decisions. The cost of education is dictated by idiosyncratic talent and the availability of mentors of the same group. Our starkest results are twofold: First, the share of minority workers in the surplus-maximizing labor force can be *higher* than in the overall population. This occurs when individual surplus varies more with talent than with mentor availability. Second, the optimal intervention in such a situation is *persistent*. It remunerates agents for their ongoing recruitment externalities on future generations.

Our model builds on the following stylized facts from the empirical literature. 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, which has a sizable impact on later compensation. Similarly, [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. The literature documents a boost in student performance and graduation rates when taught by same-group teachers ([Bettinger and Long, 2005](#); [Dee, 2007](#)). 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) or perceives other-race students as inattentive (Dee, 2005). Third, these achievement differences arise early on and manifest themselves through different education choices. For instance, the undergraduate student body for economists 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).

Formally, we consider an unsaturated, competitive labor market in continuous time that is drawing from a population composed of a majority and a minority. Every person belongs to one of two groups, which can refer to gender, race, disability status or other socio-demographic characteristics. There is a constant inflow of juniors. The instant a junior is born, he can invest in education. An educated junior seeks employment in a competitive labor market and turns into a senior. Each senior lives for an exponentially distributed time and provides mentoring to new juniors. As such, every worker provides mentoring during his entire life but receives mentoring benefits at the education stage.

The cost of education is a function of innate talent and mentoring quality. A lack of suitable mentors makes it harder for minority juniors to obtain a degree than for their peers of equal talent.<sup>1</sup> The key parameters in our model are talent concentration, mentor capacity and majority share. *Talent concentration* measures the dispersion of talent in the population. If all individuals have equal talent, the concentration is zero. High-skill sectors (doctors, lawyers, professors) have high talent concentration. We assume no ex-ante differences in talent distribution across the two groups. *Mentor capacity* captures the average number of mentees one mentor can look after at a point in time. The type of mentor interaction matters: Capacity is high for classroom instruction but low for one-on-one coaching. 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

---

<sup>1</sup>We assume a fixed mentoring technology but are agnostic about the drivers of this complementarity. Seniors from the same minority may be more efficient mentors, or they may increase acceptance and understanding for minority students more broadly.

of race in the United States, where around 76.9% of the population is white.<sup>2</sup>

Investment in education is generally inefficient since juniors only account for the mentoring they receive, not the surplus they generate for future generations. Due to competition and free entry, neither do firms internalize the effect of today's hires on tomorrow's candidate pool when setting wages. 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.<sup>3</sup> For sufficiently high mentor capacity or talent concentration, top talent from both groups works in the optimal stable steady state. Yet, the failure to internalize mentoring externalities means that even the best steady state achieves less than optimal social surplus. In those situations, a patient planner chooses to persistently intervene in favor of the minority. A surprising feature of the optimal intervention is that it may over-represent the minority group in the workforce.<sup>4</sup> This happens if the two population pools are of uneven size and mentor capacity is large.

Both temporary and persistent regulation can be implemented in different ways. We consider educational subsidies (scholarships) and workplace hiring quotas, and identify the winners and losers of each policy. In our framework, the optimal educational subsidies are budget neutral in the long run. Hiring quotas are equally effective as long as the competitive environment allows for group-specific wages. When wage disparities are restricted due to cultural norms or firm-intern politics, hiring quotas however cause significant crowding out of mediocre majority workers. The dire employment prospects dampen their investment, yet some of them still pursue an ex-post worthless education. This can result in strong opposition to

---

<sup>2</sup>See U.S. Census Bureau QuickFacts, as retrieved from <https://www.census.gov/quickfacts/fact/table/US/RHI125216> on 09/29/2017.

<sup>3</sup>This might explain why different empirical studies can reach opposing conclusions depending on whether a certain affirmative action policy has been "strong enough" to push the economy into a more desirable convergence region. It is consistent with the finding in [Bettinger and Long \(2005\)](#) that female instructors do positively influence course selection and major choice in some disciplines, but that there are no positive and significant effects in some male-dominated fields. It is also in line with the finding in [Casas-Arce and Saiz \(2015\)](#) that political parties that were most affected by a quota in Spain benefited the most in the long run.

<sup>4</sup>To be clear, the majority group still dominates in the optimal workforce, but not as much as in the population. It is never optimal for the minority to dominate in the skilled labor force, as for example in South Africa ([Commission for Employment Equity, 2017](#)). At such an (unstable) starting point, the optimal intervention at first favors the population majority.

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.<sup>5</sup>

We assume a specific functional form to present our results most intuitively, and to motivate our mentorship boost function from a discrete matching market. Our main findings are however robust to other talent distributions or mentoring technologies. We document this by showing that over-representation of the minority and persistent intervention arise without parametric assumptions as long as the marginal mentoring gain from over-representation is smaller for the majority than for the minority, and as long as surplus varies more with talent than with mentoring.

Our argument is qualitatively different from one 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 justified only insofar as it remedies historical injustice, and should render itself obsolete in a relatively short amount of 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 rare mentoring skills that do more for future talent recruitment than those of his majority twin. A surplus-maximizing intervention remunerates him for that valuable skill. It generally does so indefinitely unless the two groups make up equal shares of the underlying population. In other words, gender-based policies eventually become obsolete even under surplus maximization, but not necessarily those

---

<sup>5</sup>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), and mediocre men may be among the biggest losers.

based on race or other minority characteristics.

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.<sup>6</sup> Under statistical discrimination, employment quotas may actually reinforce negative stereotypes against certain groups (Coate and Loury, 1993). The intuition is the following: 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. 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 allowing for tangible mentoring complementarities, and show that this sheds a more positive light on affirmative action policies.

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).<sup>7</sup> Most relevant, this 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 they assume that the two population pools are of equal size, while our starkest results arise precisely when they are not. Only then do the policy recommendations go substantially beyond fairness concerns, as explained above. Only then is the optimal intervention persistent. And only

---

<sup>6</sup>Taste-based discrimination also arises when leadership is threatened by the appointment of different demographic groups. A quota can then strengthen the meritocracy of political elections, as Besley et al. (2017) empirically demonstrate.

<sup>7</sup>Becker and Tomes (1979); Restuccia and Urrutia (2004); Herskovic and Ramos (2017) apply this to overlapping generations frameworks with income and innate talent as the main input variables. They show that affirmative action is most effective if targeted towards the lower end of the income distribution. We abstract away from income differences and purely focus on cultural or gender differences that are known to play a role in mentoring relationships.

unequal pools can accurately capture a possible “glass ceiling effect” in multilevel organizations, causing us to raise doubt on some of their conjectures.<sup>8</sup> We are able to obtain these more general results and keep track of an unsaturated and decentralized labor market thanks to a novel continuous-time overlapping generations setup.

The remainder of the paper is structured as follows: In [Section 2](#) we set up our model of labor force participation and mentoring. We formally derive the evolution of an unregulated labor force in [Section 3](#), and analyze its steady states. In [Section 4](#), we define long-term social surplus and compare the optimal labor force composition between temporary and persistent policy interventions. We conclude that section by contrasting specific policy instruments under varying assumptions regarding wage determination. Finally, in [Section 5](#), we discuss the robustness of our findings with respect to the previously made parametric assumptions. [Section 6](#) concludes.

## 2 Model

### 2.1 General model

We study a game in continuous time. At every instant in time  $t \in \mathbb{R}$  two infinite pools of juniors are born, hereafter referred to as group  $i = 1$  (majority) and group  $i = 2$  (minority). Each junior is additionally indexed by the innate type  $\theta \in \Theta \equiv [0, \infty)$  which determines her talent  $x(\theta) \in [0, \infty)$ . We assume the *talent function*  $x : [0, \infty) \rightarrow (0, \infty)$  to be decreasing. The constant inflow of group- $i$  juniors with innate type  $\theta$  is given by the Lebesgue measure on  $\Theta$  scaled by a constant  $\beta_1 > \beta_2 > 0$ . More precisely, group- $i$  juniors with talent in  $(x(\theta), x(\theta')]$  arrive with constant flow rate  $\beta_i(\theta' - \theta)$  for  $\theta' > \theta$ . All qualitative results are driven by the ratio  $b = \frac{\beta_1}{\beta_1 + \beta_2} \geq 0.5$ , which we refer to as *majority share*.

Upon birth, each junior has a one-time opportunity to invest into costly education, thereafter becomes a senior, seeks employment and acts as a mentor for new juniors. Education is instantaneous, binary and the only determinant of later

---

<sup>8</sup>See [Page 22](#) for a detailed discussion, and [Page 30](#) for additional comments.

productivity.<sup>9</sup> Life expectancy follows a standard exponential distribution with parameter 1.<sup>10</sup>

There is a fixed cost of education  $c > 0$ , which is reduced both by the junior's individual talent  $x(\theta)$  and the group-specific strength of mentoring  $\mu_i$ . In an instant where there is a mass  $L_i$  of group- $i$  seniors and a total mass  $\ell$  of juniors who invest into education, the strength of mentoring  $\mu_i = \mu(\Lambda_i)$  is determined by the non-decreasing *mentorship boost function*  $\mu : [0, \infty) \rightarrow [0, 1)$ , where  $\Lambda_i = \frac{L_i}{\ell}$  is the ratio of group- $i$  seniors to juniors. This captures four important features: First, mentorship is a scarce resource that is limited by the ratio of seniors to juniors. Second, mentoring is only effective between a senior and junior of the same group. Third, if there are many juniors, it is harder for juniors to find a good match. Fourth, seniors of the opposite group then indirectly hurt juniors: They do not affect the junior's own cost of education, but attract additional juniors from the opposite group, and thereby increase competition for mentors. We discuss a micro foundation of these assumptions in [Section 2.2](#).

The benefit of education is collected once juniors become seniors. Seniors seek jobs in a competitive and unsaturated labor market. Earnings are determined through market forces: We assume that each unit mass of educated senior workers contributes one unit to a firm's profit flow, uneducated workers contribute nothing. Assuming free entry of firms, the life expectancy of 1 ensures that the expected lifetime earnings  $w_i$  of each educated senior equal 1 in an unregulated labor market, and that uneducated workers earn nothing.

Individual rationality implies that a junior invests in education if and only if her expected lifetime earnings outweigh the cost of education. Formally, a group- $i$  junior with talent  $x(\theta)$ , born in an instant with a stock  $L_i$  of group- $i$  seniors, joins

---

<sup>9</sup>Real-world examples that fit this (admittedly stylized) description include sectors where a diploma is the main hiring criterion. One may think of specialized exams such as the Bar license for lawyers or the PE license for engineers. In [Section 5](#), we discuss how the model can be adapted when other characteristics also affect productivity.

<sup>10</sup>To make sure that all strategies are defined at any instant, we assume that a senior still serves as a mentor at the time of her death.

a mass  $\ell$  of her peers in pursuing education if and only if

$$c - x(\theta) - \alpha\mu\left(\frac{L_i}{\ell}\right) \leq w_i, \quad (\text{IR})$$

where  $\alpha > 0$  measures the relative importance of mentoring versus innate talent. As in [Athey et al. \(2000\)](#), we assume that there are no complementarities between talent and mentorship boost.<sup>11</sup>

## 2.2 Talent function and mentoring boost

For expositional clarity, we use concrete parametric functions for the main part of our paper and demonstrate robustness of our core results in [Section 5](#). For talent, we consider the function

$$x(\theta) = \lambda e^{-\lambda\theta}$$

depicted in [Figure 1a](#).<sup>12</sup> This function normalizes the total talent in a unit-scale population  $\int_0^\infty \lambda e^{-\lambda\theta} d\theta$  to 1 but allows us to vary the *concentration* of talent through  $\lambda$ . High concentration means that a small elite possesses abundant talent, low concentration reduces this heterogeneity. Practically,  $\lambda$  is particularly large for specialized education that requires rare skills, such as for doctors, lawyers or actors.

For mentoring, we consider the mentorship boost function

$$\mu(\Lambda_i) = 1 - e^{-q\Lambda_i}, \quad (1)$$

where we call  $q$  the *mentor capacity*. This function can be micro founded as follows: Consider the limit of a discrete matching market with  $nL_i$  seniors of group  $i$  and  $nl$  juniors (of any group), as  $n \rightarrow \infty$ . Each senior-junior link exists with probability  $\frac{q}{nl}$ , where mentor capacity  $q$  denotes the average number of juniors a mentor can

<sup>11</sup>We discuss the relevance and implications of this assumption in [Section 5](#).

<sup>12</sup>Note that  $x$  is *not* a density function, and this is not equivalent to exponentially distributed talent. We chose this formulation to stay as close as possible to [Athey et al. \(2000\)](#), and to allow for unbounded population pools given our interest in unsaturated labor markets. [Section 5](#) considers finite population mass and makes the talent distribution explicit.

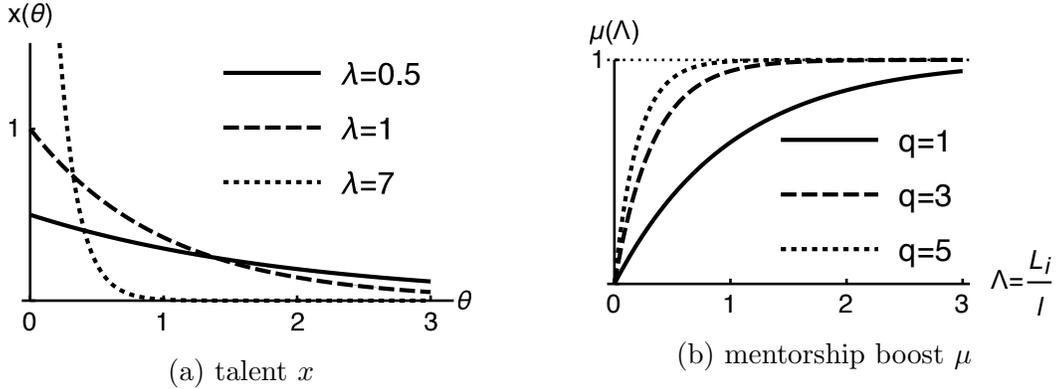


Figure 1: Parametric functions used in the main part of the paper.

oversee.<sup>13</sup> The randomness of mentor assignment captures the matching frictions mentioned above. A junior of group  $i$  enjoys a mentorship boost of 1 if and only if he or she is being mentored by at least one group- $i$  mentor.<sup>14</sup> By the Law of Rare Events, the number of same-group mentors per junior can be approximated by the Poisson distribution as  $n$  grows. The probability of finding a same-group mentor, and hence the expected mentorship boost for a junior of group  $i$ , converges to the expression in Equation (1). The mentor capacity  $q$  modifies both the level and the curvature of the mentorship boost function, as illustrated in Figure 1b. This parameter is high for industries where the relevant skills are imparted through classroom instruction, and low where mentoring requires individual coaching. Decreasing trends in the time invested in mentoring (DeLong et al., 2008) will also affect market dynamics through this channel.

Two restrictions on the cost parameter are necessary for realism and tractability, so as to ensure that labor supply never completely dries out or explodes. To this end, we impose the following parameter restriction for the remainder of the paper:

$$0 < c - 1 - \alpha\mu(1) < \lambda. \quad (\text{A1})$$

These bounds ensure that education is too costly for zero-talent juniors under the maximal steady-state mentorship boost  $\mu(1)$ , but affordable to the most talented

<sup>13</sup>For  $n$  large enough, the link probabilities are non-degenerate,  $0 < \frac{q}{nl} < 1$ .

<sup>14</sup>See Section 5 for an example where cross-group mentoring is also beneficial.

junior.

### 2.3 Equilibrium notion and steady state

In every instant, the pool of seniors  $\mathbf{L} = (L_1, L_2)$  acts as a state variable. An equilibrium is characterized by an equilibrium wage  $w_i$ , and the investment decision for each junior that satisfies the individual rationality constraints (IR). The outcome in an instant with state  $\mathbf{L}$  can be described by the mass of juniors who invest from each group  $\mathbf{l} = (l_1, l_2)$ , where  $l_i \geq 0$  denotes the mass of group- $i$  juniors who invest.

The instantaneous investment determines labor force dynamics. As educated juniors themselves turn into seniors, junior investment also affects the evolution of the senior labor force. Their inflow offsets the exit of seniors, whose individual lifespans follow a standard exponential distribution with intensity 1. Senior labor force  $\mathbf{L}(t) = (L_1(t), L_2(t)) \in \mathbb{R}_+^2$  therefore evolves according to the differential equation

$$\dot{\mathbf{L}}(t) = \mathbf{l}(t) - \mathbf{L}(t), \quad (2)$$

with  $l_i(\tau) \geq 0$  denoting the mass of group- $i$  juniors born at time  $\tau$  who invest in education given the current stock of seniors being  $L(t)$ .

Besides the intertemporal dynamics of the labor force, we are interested in long run steady state outcomes. In line with the literature, we define steady states as fixed points of the system and stable steady states as fixed points that are robust to small perturbations.<sup>15</sup>

**Definition.** *The economy is in a steady state when  $\dot{\mathbf{L}} = 0$  or, equivalently,  $\mathbf{l} = \mathbf{L}$ . A steady state  $\hat{\mathbf{L}} \in \mathbb{R}_+^2$  is stable whenever a small perturbation does not affect the long-term convergence, i.e., when there exists  $\varepsilon > 0$  such that for all  $\mathbf{L}(0) \in \mathbb{R}_+^2$  with  $\|\mathbf{L}(0) - \hat{\mathbf{L}}\| < \varepsilon$ ,  $\lim_{t \rightarrow \infty} \mathbf{L}(t) = \hat{\mathbf{L}}$ .*

For brevity of exposition, we drop all dependence on  $t$  for the remainder of this paper and phrase everything in terms of the two-dimensional state variable  $\mathbf{L}$ . This is possible because junior investment (IR) depends on calendar time only

---

<sup>15</sup>This same concept is sometimes referred to as “asymptotically stable” in the literature.

through the senior labor force  $\mathbf{L}$ . Exponential life expectancy (2) ensures the same for labor force evolution  $\dot{\mathbf{L}}$ .

### 3 Dynamics of an unregulated economy

#### 3.1 Uniqueness of the instantaneous equilibrium

Combining individual rationality constraints (IR) across juniors, the equilibrium conditions for junior investment  $\mathbf{l} = (l_1, l_2)$  given the mentoring pool  $\mathbf{L} = (L_1, L_2)$  are

$$\begin{cases} c - x \left( \frac{l_1}{\beta_1} \right) - \alpha\mu \left( \frac{L_1}{l_1+l_2} \right) = w_1 \text{ or } \left( l_1 = 0 \text{ and } c - \lambda - \alpha\mu \left( \frac{L_1}{l_1+l_2} \right) > w_1 \right), \\ c - x \left( \frac{l_2}{\beta_2} \right) - \alpha\mu \left( \frac{L_2}{l_1+l_2} \right) = w_2 \text{ or } \left( l_2 = 0 \text{ and } c - \lambda - \alpha\mu \left( \frac{L_2}{l_1+l_2} \right) > w_2 \right), \end{cases} \quad (3)$$

where wage  $w_i = 1$  for the case without market intervention. This system of equations admits a unique solution  $\mathbf{l}$  that responds predictably to changes in wage. (In Section 4.4, we discuss how quotas may generate group-specific wages.)

**Proposition 1.** *The equilibrium investment by juniors,  $\mathbf{l} \in \mathbb{R}_+^2 = [0, \infty)^2 \setminus \{(0, 0)\}$ , is uniquely determined by (3) for any senior labor force  $\mathbf{L} \in \mathbb{R}_+^2$ . Moreover, group- $i$  investment  $l_i$  is increasing in  $w_i$  and decreasing in  $w_j$  for  $j \neq i$ .*

*Proof.* See Appendix A. □

#### 3.2 Steady state analysis

In this section, we identify necessary and sufficient condition for the existence of both homogeneous and mixed stable steady states. In doing so, we are primarily interested in the labor force *composition*  $\phi = \frac{L_1}{L_1+L_2}$  rather than its total *size*  $L = L_1 + L_2$ .<sup>16</sup> We will show that the mixed stable steady state with composition  $\hat{\phi}$  generally over-represents the majority relative to the population,  $\hat{\phi} > b$ . Some of

---

<sup>16</sup>Indeed, the former is a sufficient for the steady state analysis since Equation (3) uniquely determines the corresponding  $L$  (see Lemma 2 in the appendix).

our results involve limits, and so we strengthen [Assumption \(A1\)](#) to

$$c - 1 - \alpha > 0. \tag{A1^+}$$

In other words, we assume that educational investment remains bounded even under the maximal mentorship boost  $\mu = 1$ .

**Theorem 1** (Steady States). *Consider an economy that satisfies [Assumption \(A1\)](#).*

(a) *The economy admits two homogeneous steady states  $\phi \in \{0, 1\}$  if and only if*

$$c - \lambda \geq 1. \tag{hSS}$$

*They are stable if and only if the inequality [\(hSS\)](#) is strict.*

(b) *The economy admits a mixed steady state  $\phi \in (0, 1)$  if and only if*

$$c - 1 - \alpha\mu(0.5) < x(0) = \lambda. \tag{mSS}$$

(c) *Under [Assumption \(A1<sup>+</sup>\)](#) and for sufficiently high mentor capacity  $q$  or talent concentration  $\lambda$ , there exists a unique stable mixed steady state with  $\phi$  arbitrarily close to majority share  $b$ .*

(d) *Any stable mixed steady state over-represents the majority,  $\hat{\phi} > b$ , whenever  $\beta_1 > \beta_2$  and*

$$c - 1 - \alpha > \alpha e^{-q/2}. \tag{mSS^+}$$

*Proof.* See [Appendix A](#). □

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 completely once it is severely underrepresented in the workforce. [Property \(mSS\)](#) states that mentor availability  $\Lambda = 0.5$  is sufficient to attract at least *some* investment. This is of course necessary for the

existence of a mixed steady state (claim b) where at most one group may exceed the mentorship boost  $\mu(0.5)$ . As it turns out, it is also sufficient.

The economy may admit multiple stable mixed steady states, but we can offer a partial characterization of their composition. To gain intuition, consider the break-even talent  $\hat{x}_i$  that solves Equation (IR). If either mentor capacity or talent concentration is large,  $\hat{x}_i$  hardly responds to differences in mentor availability since  $\mu(\Lambda) \rightarrow 1$  as  $q \rightarrow \infty$  and  $x'(x^{-1}(\hat{x})) \rightarrow -\infty$  as  $\lambda \rightarrow \infty$ . As a result, the mixed stable steady state is unique and approaches majority share  $b$  (claim c). More generally, we show that *all* stable mixed steady states involve an over-representation of the majority under (mSS<sup>+</sup>) due to mentoring frictions (claim d). Property (mSS<sup>+</sup>) can be interpreted as a lower bound on  $c$ ,  $q$  or  $\alpha$ .<sup>17</sup>

To prove Theorem 1 formally, we rely on the Hartman-Grobman Linearization Theorem to characterize the steady states of the labor force. In a visual representation of the labor force evolution  $\dot{\mathbf{L}}$  as a vector map, the theorem formalizes the idea that a steady state is stable if and only if all surrounding arrows point towards it. In the case of Figure 2,<sup>18</sup> there are three stable steady states: Two are *homogeneous*  $(L_1, 0), (0, L_2) \in \mathbb{R}_+^2$  and one is *mixed*. The starting pool of mentors  $\mathbf{L} \in \mathbb{R}_+^2$  determines towards which of these the labor force will ultimately converge.

**Theorem** (Hartman-Grobman Linearization Theorem). *A steady state  $\hat{\mathbf{L}}$  is stable if and only if all the eigenvalues of the Jacobian matrix  $\left. \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}} \right|_{\mathbf{L}=\hat{\mathbf{L}}}$  have a negative real part.*

To establish the results in Theorem 1, we reduce the dimensionality of the problem by mapping all (stable) steady state compositions to the (downward crossing) zeros of a one-dimensional auxiliary function. A convexity argument then rules out downward crossing zeros over  $(0, b]$ . The following lemma captures the core of the proof.

---

<sup>17</sup>The properties and assumptions are not mutually exclusive. Indeed, they hold for any education cost  $c > 1$ , mentoring importance  $\alpha < c - 1$  (A1<sup>+</sup>), talent concentration  $c - 1 - \alpha < \lambda \leq c - 1$  (hSS), and mentor capacity  $q$  large enough to satisfy (mSS) and (mSS<sup>+</sup>).

<sup>18</sup>Unless otherwise indicated, all figures are obtained with parameter values  $c = 2.1$ ,  $\alpha = 0.6$ ,  $\lambda = 1$ ,  $\beta_1 = 1.2$ ,  $\beta_2 = 1$  and  $q = 3$ .

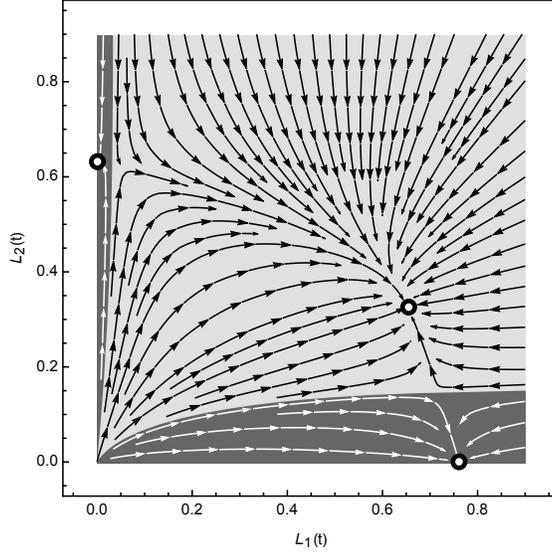


Figure 2: Labor force evolution  $\dot{\mathbf{L}}$  at any  $\mathbf{L}$ . The three stable steady states are highlighted along with their respective convergence regions.

**Lemma 1.** Let  $S : [0, 1] \rightarrow \mathbb{R}$  be given by

$$\begin{aligned}
 S(\phi) &= (1 - \phi)bx^{-1}(c - 1 - \alpha\mu(\phi)) - \phi(1 - b)x^{-1}(c - 1 - \alpha\mu(1 - \phi)) \quad (4) \\
 &= \frac{1}{\lambda} \left[ (b - \phi) \ln(\lambda) - (1 - \phi)b \ln(c - 1 - \alpha\mu(\phi)) \right. \\
 &\quad \left. + \phi(1 - b) \ln(c - 1 - \alpha\mu(1 - \phi)) \right].
 \end{aligned}$$

Under *Property (mSS)*, there exists a mixed steady state of composition  $\phi$  if and only if  $S(\phi) = 0$ . The steady state is stable if and only if  $S'(\phi) < 0$ .

*Proof.* *Property (mSS)* ensures that total labor supply  $L$  is positive. The one-to-one correspondence between steady-state composition and the roots of  $S$  then stems directly from [Equation \(3\)](#) at a mixed steady state  $\mathbf{L} = \mathbf{l} = (\phi L, (1 - \phi)L)$ , since

$$\begin{aligned}
 (1 - \phi)bx^{-1}(c - 1 - a\mu(\phi)) &\stackrel{(3)}{=} \frac{\phi(1 - \phi)}{\beta_1 + \beta_2} L \stackrel{(3)}{=} \phi(1 - b)x^{-1}(c - 1 - a\mu(1 - \phi)) \\
 &\Leftrightarrow S(\phi) = 0.
 \end{aligned}$$

Stability owes to the Hartman-Grobman Linearization Theorem and is relegated to [Appendix A](#).  $\square$

In addition to these existence results, [Lemma 1](#) also allows us to anticipate the impact of parameter changes on the labor composition in a mixed stable steady state. [Figure 3](#) illustrates the result for changes in mentor capacity  $q$ .

**Theorem 2** (Comparative Statics). *Consider an economy that admits a mixed stable steady state  $\phi \in (b, 1)$  over unequal talent pools ( $b > 0.5$ ). As*

(a) *pool sizes become more even,  $\tilde{b} \in (0.5, b)$ ,*

(b) *talent becomes more concentrated,  $\tilde{\lambda} > \lambda$ , or*

(c) *mentor capacity increases,  $\tilde{q} > q$ ,*

*there exists a mixed stable steady state  $\tilde{\phi}$  with smaller group-imbalance  $\tilde{\phi} \in (0.5, \phi)$ .*

*Proof.* See [Appendix A](#).  $\square$

It is not surprising that the composition of the population (captured by majority share  $b$ ) is positively related to that of the steady state labor force. What is more interesting is that the strength of this relation varies across sectors. First, higher talent concentration  $\lambda$  means that innate talent is more important to be successful. Thus, conditional on being in a mixed steady state, we should expect a more balanced labor force in high-skill professions. Similarly, skills taught through class room instruction (for example college education) generate a lower steady-state group imbalance than those acquired through individual advising (during e.g. graduate school or executive coaching) where mentor capacity  $q$  is smaller.

To illustrate how to interpret these results, let us consider the “glass ceiling effect”: It is a well-known phenomenon that the higher up the career ladder, the larger is the realized group-imbalance.<sup>19</sup> Theorems 1 and 2 suggest that this can be due to two reasons. Either, we might be moving from a homogenous steady state

---

<sup>19</sup>[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.

to a mixed steady state due to a historic shock, and it takes several generations to converge to the new steady state composition. Or, the echelons of the career ladder are fundamentally different jobs that require different skills. Lower-ranked jobs require skills that are less concentrated in the population, but are taught through classroom instruction where mentor capacity is high. If group-imbalance is larger at the top in the steady state, this would imply that the reduction in mentor capacity outweighs the increase in talent concentration, at least from a positive perspective. In the next section, we study the related normative question of what the composition *should be* in order to maximize total surplus.

## 4 Optimal Policy Intervention

### 4.1 Welfare metric

Mentoring complementarities generate a tension between talent recruitment and mentoring efficiency: Only a homogeneous labor force ensures perfect within-group mentor assignments, but a more balanced labor force harnesses the top talent from both groups. We base our efficiency analysis on total surplus, measured as total productivity net educational investments,

$$\begin{aligned} \pi(\mathbf{L}, \mathbf{l}) &= \sum_{i=1}^2 \int_0^{l_i} 1 - c + x(\theta/\beta_i) + \alpha\mu \left( \frac{L_i}{l_1 + l_2} \right) d\theta \\ &= \sum_{i=1}^2 \beta_i (1 - e^{-\lambda \frac{l_i}{\beta_i}}) + l_i \left( 1 - c + \alpha\mu \left( \frac{L_i}{l_1 + l_2} \right) \right), \end{aligned} \quad (5)$$

where  $\mathbf{L} \in \mathbb{R}_+^2$  denotes the senior labor force and  $\mathbf{l} \in \mathbb{R}_+^2$  the new recruits. 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.

### 4.2 Optimal steady state

In a first step, we provide policy recommendations for interventions that are limited in time. The temporary nature of the intervention allows us to restrict our analysis

to (stable) steady states, as no other labor force can be sustained in the long term absent ongoing market intervention. This simplifies the computation of surplus. Indeed, when the labor force is constant ( $\mathbf{L} = \mathbf{l}$ ), total surplus can be written as

$$\tilde{\pi}(\phi, L) = \pi((\phi L, (1 - \phi)L), (\phi L, (1 - \phi)L)) \quad (6)$$

for labor force *size*  $L = L_1 + L_2$  and *composition*  $\phi = L_1/L$ .

We have already established that there are often multiple stable steady states ([Theorem 1](#)). A patient social planner cares to know which one maximizes surplus, so that he can redirect the economy through a temporary intervention.

**Theorem 3** (Optimal Steady State). *For sufficiently large mentor capacity  $q$  or high talent concentration  $\lambda$ , the surplus-maximizing stable steady state is mixed,  $\phi_{SS}^* \in (0, 1)$ .*

*Proof.* See [Appendix A](#). □

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 mixed steady state. Similarly, if talent is sufficiently concentrated, the extra surplus from the most able individuals outweighs any possible mentoring losses. In other words, temporary market intervention is warranted when minority participation rates threaten to vanish in an industry where talent is highly concentrated or mentoring is sufficiently broad. This makes high-skill sectors with mentoring through classroom instruction (such as graduate education) prime candidates for temporary course correction in favor of the underrepresented group.

Combining the insight from [Theorems 1](#) and [3](#) yields another important take-away: Temporary intervention does not achieve a workforce that accurately reflects the diversity in the population. Indeed, under [Property \(mSS<sup>+</sup>\)](#), the minority (with  $\beta_2 < \beta_1$ ) remains *underrepresented* at the mixed steady state in the sense that  $\phi > b$ . This has to do with the fact that minority mentors are harder to come by, making it impossible to sustain proportional participation without ongoing intervention. Perhaps surprisingly, we now show that an optimal long-term policy

often *overrepresents* the minority for this precise reason.

### 4.3 Optimal long-run intervention

For tractability, we focus our analysis on the long-run surplus from *constant* interventions by maximizing  $\tilde{\pi}(\phi, L)$ , rather than characterizing the most efficient path  $\dot{\mathbf{L}}^*(\mathbf{L})$ .<sup>20</sup> This is relevant for a patient social planner who cannot adjust his diversity targets over time. We are, at this point, agnostic about the exact implementation of the policy goal. We simply assume that the planner can directly choose *any* labor force composition  $\phi$  and total participation  $L$ . In Section 4.4, we show that such a goal can indeed be implemented through educational scholarships or hiring quotas (as long as market wages are unrestricted).

Our main result is that the welfare-maximizing labor force coincides with a mixed steady state *if and only if* the two pools are of the same size. In all other cases, it is generally optimal for a social planner to intervene *persistently*. The optimal labor force weighs talent recruitment against mentor assortativity. When mentoring is indispensable for participation (hSS) but capacity is small, a homogeneous labor force is most efficient. Larger mentoring capacities make mentoring mismatch less costly, since a junior is generally advised by multiple mentors. At some point, the optimal labor force actually *overrepresents* minority worker relative to the population. This advocates recruitment of minority workers with talent *below* the marginal majority worker – not just as a transitory course correction, but as an ongoing policy. The stark result has a simple intuition: Students 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 mentoring advantage of the majority.

**Theorem 4** (Optimal Intervention). *Long-run surplus  $\tilde{\pi}(\phi, L)$  is maximized at some  $\phi^* \in [0.5, 1)$  and  $L^* > 0$ . Moreover, the optimal labor-force composition  $\phi^*$  depends on mentor capacity  $q$  and talent concentration  $\lambda$ :*

---

<sup>20</sup>We should note that it is entirely possible to numerically describe the surplus-maximizing path based on the partial differential equation  $0 = (1 + r)\pi_{\mathbf{l}}(\mathbf{L}, \mathbf{l}) + \pi_{\mathbf{L}}(\mathbf{L}, \mathbf{l}) + \pi_{\mathbf{L}\mathbf{l}}(\mathbf{L}, \mathbf{l})(\mathbf{L} - \mathbf{l})$ , where  $r$  is the planner’s discount rate. Our optimal intervention corresponds to the steady state of this dynamic policy as  $r \rightarrow 0$ .

- (a) If *Property (hSS)* holds with strict inequality and  $q$  is small enough, the optimal labor force is homogeneous,  $\phi^* = 1$ .
- (b) If  $q$  is large enough, the optimal labor force over-represents the minority  $\phi^* \in [0.5, b]$ . The bounds are strict  $\phi^* \in (0.5, b)$  whenever  $\beta_1 > \beta_2$ .
- (c) Under  $(A1^+)$ , the optimal composition converges to that of the population  $\lim_{q \rightarrow \infty} \phi^* = b$ .
- (d) For large enough  $\lambda$ , the optimal labor force is more balanced than in the unique mixed stable steady state,  $0.5 \leq \phi^* < \hat{\phi}$ , if and only if  $q > Q(b) = \frac{1}{2b-1} \ln \left( \frac{b}{1-b} \right)$ .

Whenever the composition  $\phi^*$  corresponds to a steady state, so does total participation  $L^*$ .

*Proof.* See [Appendix A](#). □

To illustrate the individual claims, [Figure 3](#) plots the optimal and steady state labor force composition against mentor capacity. A stable mixed steady state exists for  $q \geq 2.74$ ; it involves an over-representation of the majority. For small  $q < 2.76$ , the surplus-maximizing labor force completely excludes the minority (claim a). For intermediate  $q$ , the minority participates but is under-represented relative to the population ( $\phi \in (b, 1)$ ). For large  $q > 4$ , this is reversed: The optimal minority share in the workforce exceeds that of the population (claim b). As  $q$  grows further, optimal and steady-state composition move towards that of the population (claim c). Even for moderate mentor capacity, the optimal labor force is generally more diverse than the steady state composition, as long as talent is sufficiently concentrated (claim d). As a point of reference, we mention in the introduction that the majority share of whites in the US is at roughly  $b = 0.77$ . If the average ratio of students per mentor exceeds  $Q(0.77) = 2.24$ , ongoing policies in favor of racial minorities may therefore enhance efficiency in high-skill sectors.

To explain the intuition, [Figure 4](#) displays the social surplus visually, as the area under the talent curve up to the break-even point  $\hat{x}_i = c - 1 - \alpha\mu(\Lambda_i)$  for both majority (hatched) and minority (solid gray). Decreasing returns-to-scale from

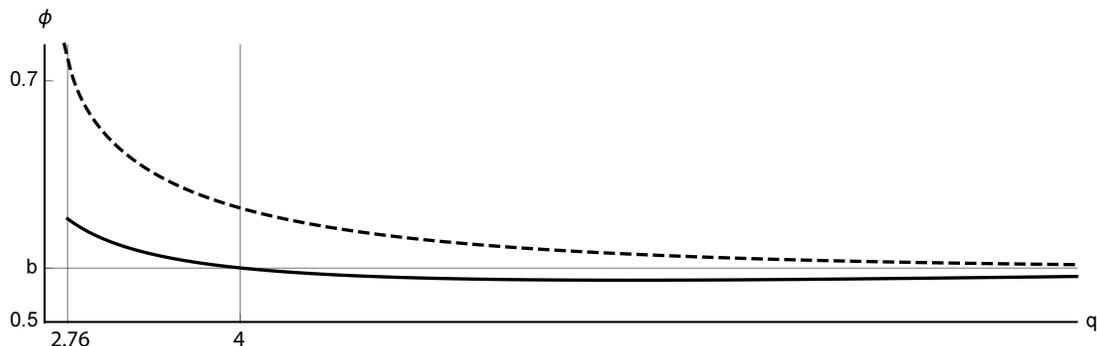


Figure 3: Optimal (solid) and steady state (dashed) labor force composition as a function of mentor capacity.

mentoring ensure that moving to a more balanced workforce affects the break-even point for the minority more strongly,  $\mu'(1 - \phi) \geq \mu'(\phi)$  for  $\phi > 0.5$ . This by itself is not enough to raise overall surplus however, since the smaller  $\mu'(\phi)$  affects a bigger mass of mentees  $\phi$ . Visually, the question is not whether the vertical change in  $\hat{x}_1$  dominates that in  $\hat{x}_2$ , but which of the (green or blue) areas dominates. Of course, when the cutoff  $\hat{x}_1$  hardly responds at all ( $q \rightarrow \infty$ ), the former will eventually imply the latter. Alternatively, when talent is highly concentrated ( $\lambda \rightarrow \infty$ ), the colored areas approach rectangles of width  $\phi L$  and  $(1 - \phi)L$  respectively. It is now possible to establish a condition of ‘sufficiently decreasing returns to scale’ (captured by  $Q(b)$ ) that dictates the direction of the change in surplus. Loosely speaking, persistent intervention in favor of the minority is beneficial when individual surplus varies more with talent than with mentor availability.

Together, the conclusions of [Theorems 1](#) and [4](#) reach beyond the special case of constant labor provision: They imply that a sufficiently patient planner intervenes persistently in favor of the minority in industries where the mentor-to-mentee ratio is high enough, talent is concentrated, and the two pools are of unequal size.<sup>21</sup> In particular, there is no reason to assume that affirmative action policies render themselves obsolete by virtue of their own success, contrary to the 2003 Supreme Court ruling which argued that “race-conscious admissions policies must

<sup>21</sup>Indeed, the result shows one particular intervention that – while not necessarily fully optimal – generates positive social surplus, and hence dominates a ‘laissez-faire’ regime.

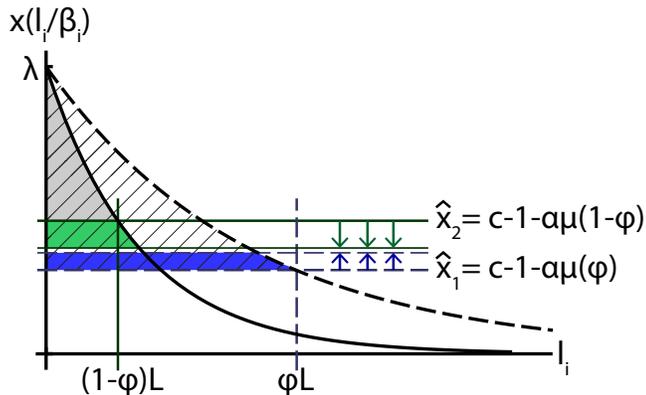


Figure 4: Changes in long-run social surplus when  $\beta_1 + \beta_2 = 1$ .

be limited in time” and expected them to disappear within 25 years.<sup>22</sup> [Theorem 4](#) also points to differences between race- and gender-based affirmative action. It suggests a larger scope for welfare gains when the two pools are of uneven size. In other words, we expect gender-based policies to be necessary only in the short run (since  $b \approx 0.5$ ), but see grounds for ongoing race-based policies (since  $b \gg 0.5$ ), particularly in high-skill sectors.

It is useful to contrast our results with [Athey et al. \(2000\)](#)’s conjecture regarding the “glass ceiling”. In their model, senior management plays the role of a surplus-maximizing social planner; which is a different angle than the decentralized view we have taken on [Page 16](#). They 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.<sup>23</sup> Our analysis warrants a more nuanced view: (a) While a population increase shifts the optimal workforce

<sup>22</sup>Grutter v. Bollinger, 539 U.S. 306 (2003), pages 309-310.

<sup>23</sup>In their words: “If the initial ability of one type becomes less scarce, the promotion rates and bias shift in favor of that type. One implication is that the entry of more women and minorities into the workforce will cause firms to shift their optimal bias toward these groups. Another implication is that in a multilevel organization, diversity might fall in higher levels of the hierarchy. This follows because inequities in one level will lead to lower promotion rates for the disadvantaged type, and thus inequities at one level will be reinforced at the next-higher level.” ([Athey et al., 2000](#), p.778f)

representation towards that group, the *bias* (which they define as a lower talent threshold) may actually be in favor of the other group, and (b) faced with an uneven middle management, optimal promotion decisions at the top may very well over-represent the dominated group.<sup>24</sup> Thus, mentoring frictions alone do not persuasively explain increasing attrition across echelons of the career ladder.

## 4.4 Policy Instruments

In the previous sections, we have shown conditions under which the policy maker wishes to ensure junior investment  $l^*$  rather than the myopic  $l$  given senior labor force  $\mathbf{L}$  with  $L_1 > L_2$ . We now turn our focus to the practical implementation of such a policy. We compare two methods that can be expressed within our simple model: Group-specific tuition schedules versus hiring restrictions. We first discuss the implications in the text and then summarize the formal results in Theorem 5.

**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.<sup>25</sup> Let  $\Delta \in \mathbb{R}^2$  denote such a transfer schedule where  $\Delta_i$  represents the net transfer to group  $i$ . Because the labor market remains unrestricted, expected returns to education remain equal to  $w = 1$ . Consequently, equilibrium investment  $l^*$  under  $\Delta$  satisfies

$$c + \Delta_i - x \left( \frac{l_i^*}{\beta_i} \right) - \alpha \mu \left( \frac{L_i}{l_1^* + l_2^*} \right) = 1 \quad \text{for } i = 1, 2. \quad (7)$$

We now show that the surplus-maximizing labor force can be implemented in a way that approaches budget balance. Indeed, the surplus-maximizing labor force

---

<sup>24</sup>Moreover, if promotions represent the top end of the talent distribution, there is no reason to assume any interactions between levels at all.

<sup>25</sup>Whenever we discuss fellowships  $\Delta_i > 0$ , those are assumed to be available to *all* interested minority students. It is straightforward to see that ability-based fellowships *only* affect the extensive margin if the available pool exceeds the unregulated student supply obtained from Equation (3). This may explain why studies such as Prenovitz et al. (2016) fail to observe additional minority recruitment for competitive scholarship programs on a very limited budget.

satisfies the first order condition

$$\begin{aligned}
0 &= \frac{\partial \tilde{\pi}}{\partial L}(\phi^*, L^*) \\
&= 1 - c + \phi^* x \left( \frac{\phi^* L^*}{\beta_1} \right) + (1 - \phi^*) x \left( \frac{(1 - \phi^*) L^*}{\beta_2} \right) \\
&\quad + \alpha \phi^* \mu(\phi^*) + \alpha(1 - \phi^*) \mu(1 - \phi^*).
\end{aligned} \tag{8}$$

Comparing [Equations \(7\) and \(8\)](#), we observe that the net investment disappears as  $\mathbf{L}$  tends to  $(\phi^* L^*, (1 - \phi^*) L^*)$  since  $l_1^* \Delta_1 + l_2^* \Delta_2 \rightarrow L \frac{\partial \tilde{\pi}}{\partial L}(\phi^*, L^*) = 0$ .

**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.<sup>26</sup> We call a quota  $\phi^*$  *binding* at  $\mathbf{L}$  if it forces the firm to recruit more minority members than they would myopically. Formally, if  $\mathbf{l}$  denotes the solution to [Equation \(3\)](#) under wages  $w = 1$ ,  $\phi^*$  is binding if and only if  $\phi^* < \frac{l_1}{l_1 + l_2}$ .

Firm competition ultimately determines the market wage and 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* group- $i$  individuals by  $\bar{l} \geq \mathbf{l}^*$  respectively. We use  $\tilde{w}_i$  to denote the flow of wage payments to employed group- $i$  individuals, resulting in expected lifetime earnings of  $w_i = l_i^* / \bar{l}_i \cdot \tilde{w}_i$  for each educated group- $i$  job candidate.

There are limits to the compositions that can be implemented through a quota.

---

<sup>26</sup>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.

Notably, the zero-profit condition for firms equates the marginal cost of an educated hire to the added productivity,

$$\phi^* \tilde{w}_1 + (1 - \phi^*) \tilde{w}_2 = 1, \quad (\text{MC1})$$

Quotas generate desired investment levels only if educational investment is sufficiently responsive to wage differences.<sup>27</sup> To avoid lengthy digressions, we study situations where education retains a positive net cost even to the most able individual under perfect mentoring,

$$c - \lambda - \alpha > 0. \quad (\text{A2})$$

Note that if  $\alpha \leq 1$  or  $\lambda < 1$ , this follows from [Property \(hSS\)](#) or [\(A1<sup>+</sup>\)](#) respectively. We also assume that the senior labor force contains at least *some* minority mentors,  $L_2 > 0$ , to avoid the possibility of a complete labor market shutdown.<sup>28</sup>

When wages are determined solely through market forces, all educated workers find employment,

$$\bar{l} = l^*. \quad (\text{MC2a})$$

Indeed, any oversupply of educated group- $i$  workers would drive their lifetime earnings to zero. Anticipating this, none of them would invest into education under [Assumption \(A2\)](#), creating a contradiction. The size of the cohort  $\bar{l}$  is uniquely determined by the market wages  $w_i$  via the individual rationality constraints ([Proposition 1](#)). Taken together, the market clearing conditions [\(MC1\)](#) and [\(MC2a\)](#) thus imply that a binding quota *raises* minority and *depresses* majority earnings relative to the unconstrained market,  $\tilde{w}_1 = w_1 < 1 < \tilde{w}_2 = w_2$ . Contrary to scholarships, a quota delegates the decision over the *size* of new hires to myopic firms and only imposes bounds on their *composition*. For the constant intervention studied in [Theorem 4](#), this is however without long-term efficiency loss as firms and planner

---

<sup>27</sup>To avoid anticipatory investment, we assume that quotas are unexpected.

<sup>28</sup>Without minority mentors and if mentoring is critical ( $\alpha \gg 0$ ), a quota shuts down the labor market: Minority investment remains zero even under the largest possible wages and consequently, firms can hire nobody. In contrast, for any  $L_2 > 0$ , [Assumption \(A1\)](#) ensures that any wage  $\tilde{w}_2 \geq 1$  attracts minority students as long as majority investment  $l_1$  is small enough. Low wages  $\tilde{w}_1 \rightarrow 0$  have the desired effect by [Assumption \(A2\)](#).

agree on the optimal total participation. Indeed, the zero-profit condition (MC1) is equal to budget balance when hiring bonuses are restated as subsidies  $\Delta = \mathbf{1} - \mathbf{w}$ .

In some industries however, social or legal pressure prohibits paying unequal wage to employees in the same position.<sup>29</sup> Without differential hiring bonuses,

$$\tilde{w}_1 = \tilde{w}_2 \tag{MC2b}$$

and the zero-profit condition (MC1) forces these market wages to 1. However, a binding quota caps the demand for group-1 workers at  $\frac{\phi^*}{1-\phi^*}\bar{l}_2 < 1$ , while all  $\bar{l}_2$  educated minority workers are hired. Workers factor this employment insecurity into their cost-benefit analysis of education (3) by expecting lifetime earnings of  $w_1 = \frac{\phi^*}{1-\phi^*}\frac{\bar{l}_2}{\bar{l}_1}$  and  $w_2 = 1$ . Relative to the unregulated economy, this implies a drop in total investment  $\bar{l}$ , as well as equilibrium *over-investment* by the majority. The reason is simply that a more balanced student body requires reduced lifetime earnings for the majority by Proposition 1. 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.

Theorem 5 summarizes the results from this section.

**Theorem 5** (Policy Instruments). *The policy maker can implement the optimal long-run labor force  $\mathbf{L}^*$  through educational subsidies that are budget neutral in the long run. Under Assumption (A2), a hiring quota  $\phi^* = \frac{L_1^*}{L_1^* + L_2^*}$  implements the same  $\mathbf{L}^*$  if and only if group-specific hiring bonuses are feasible. Otherwise, the quota reduces total investment, and causes a positive mass of majority workers to invest into education yet fail to secure employment.*

*Proof.* See text and footnotes. □

To illustrate, Figure 5 plots the evolution of key labor market variables under a hiring quota. The optimal long-run composition  $\phi^*$  is more diverse than the

---

<sup>29</sup>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](http://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).

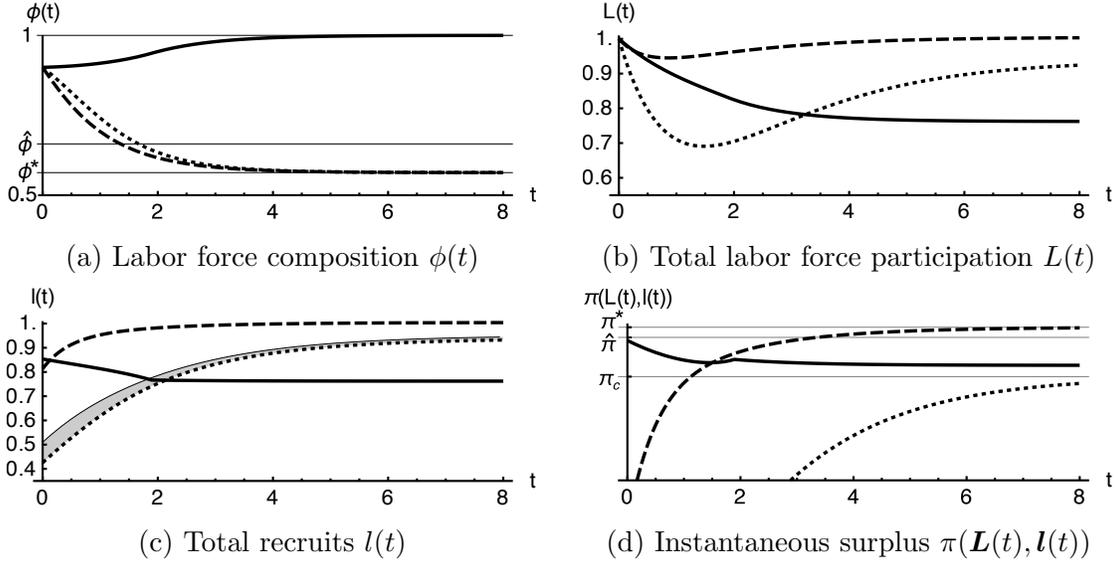


Figure 5: Labor force evolution starting at  $\mathbf{L} = (0.8, 0.2)$  without market intervention (solid line) or under a quota that imposes the optimal long-run composition  $\phi^* = 0.57$  in an environment with flexible (dashed) or common (dotted) wages.

mixed steady state  $\hat{\phi} > \phi^*$ . The starting value  $\mathbf{L}(0)$  is such that an unconstrained economy (solid line) converges to a homogeneous steady state (panel a). Imposing quota  $\phi^*$  under flexible (dashed) or common wages (dotted) causes a temporary drop in total labor force participation (panel b) and investment (panel c). The effect is more pronounced under common wages where the quota amounts to a hiring cap on majority workers. This leads to educational over-investment (shaded area) from majority workers who ultimately remain unemployed. Costly over-investment persists in the long run; causing surplus to converge to a level  $\pi_C$  well below the optimal level  $\pi^*$ , and even below the *laissez-faire* regime (panel d). In such an environment, the policy maker could use scholarships to recover the dashed path. Alternatively, he can implement a temporary quota that merely redirect the economy towards the most efficient steady state  $\hat{\phi}$ , yielding long-term surplus  $\hat{\pi}$ .

## 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 boosts the lifetime productivity of a worker with talent  $x$  and mentorship boost  $\mu$  by  $\omega_1 x + \omega_2 \mu + \omega_3$ . 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  $c \mapsto 1 + \frac{c - \omega_3}{1 + \omega_1}$  and  $\alpha \mapsto \frac{\alpha + \omega_2}{1 + \omega_1}$  can then map this situation into our existing model. Indeed, the mapping transforms the previously assumed individual surplus

$$1 - (c - x - \alpha\mu) \mapsto \frac{1}{1 + \omega_1} \cdot \left[ \underbrace{\omega_1 x + \omega_2 \mu + \omega_3}_{\text{productivity}} - \underbrace{(c - x - \alpha\mu)}_{\text{cost}} \right]$$

into a constant fraction of the new individual surplus. 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.

**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 path-dependence that results from any such interaction or history-dependence in mentoring and talent. For tractability of the steady state analysis, we both rely on the sufficiency of the current workforce for its future evolution. 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 ‘attune’ 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.

**Nonparametric distributions.** To demonstrate the robustness of our main findings, we here abstain from parametric assumptions regarding talent distribution and mentorship boost. In line with the discussed above, we maintain the additive structure by setting education cost equal to  $c - x_i - \alpha\mu\left(\frac{L_i}{l_i}\right)$  for a group- $i$  junior of talent  $x_i$ . However, we allow for any absolutely continuous talent distribution  $F$  over  $\mathbb{R}$ , and any twice differentiable, increasing mentorship boost function  $\mu : [0, \infty) \rightarrow [0, 1]$ . To incorporate different pool sizes, we assume that there is a mass  $\beta_i(1 - F(x))$  of group- $i$  individuals of ability at least  $x$ .<sup>30</sup>

Individual optimality conditions (IR), labor market evolution and welfare are as before, save a change of variables from rank  $\theta$  to ability level  $x$ . Total surplus (5) then equals

$$\pi(\mathbf{L}, \mathbf{l}) = \sum_{i=1}^2 \int_{F^{-1}\left(1 - \frac{l_i}{\beta_i}\right)}^{\infty} \left(1 - c + x + \alpha\mu\left(\frac{L_i}{l_1 + l_2}\right)\right) \beta_i F'(x) dx. \quad (9)$$

For tractability, we assume that  $0 < F(c - 1) < 1$ ; some but not all students invest into education even without mentoring.<sup>31</sup> We also assume that  $F$  is strictly increasing over its convex range, so that there exist individuals at each intermediate ability level. Finally, we assume that  $\mu'$  and  $\mu''$  are both bounded over the interval

---

<sup>30</sup>Because our talent function  $x$  assumes infinite population pools, there is no direct translation between the two frameworks. In spirit,  $F$  is the closest analogue to  $1 - x^{-1}(x)$ . The slight switch in assumptions is deliberate: It allows to demonstrate in particular that our findings extend to normally distributed talent.

<sup>31</sup>This assumption violates [Property \(hSS\)](#), and thus rules out completely homogeneous steady states. However, the restriction still allows for multiple steady states, including almost homogeneous ones. The assumption trivially holds whenever ability is unbounded above and below, as in the case of a normal distribution.

[0, 1].

The main results of our paper are two-fold: The optimal labor force composition *over-represents* minorities relative to their population share, and a patient planner adopts a persistent, rather than temporary, market intervention. Both results generalize to this broader class of functions, as long as the marginal mentoring gain from over-representation is smaller for the majority than it is for the minority,

$$\mu(b) + b\mu'(b) < \mu(1 - b) + (1 - b)\mu'(1 - b) \quad (10)$$

and skill recruitment dominates mentoring ( $\alpha$  small). This echoes the conditions in [Theorem 4\(d\)](#): Mentoring must have sufficiently decreasing returns-to-scale to satisfy (10), and individual surplus is primarily driven by talent rather than mentoring.

**Theorem 6** (Nonparametric). *Suppose  $\beta_1 > \beta_2$ . If [Property \(10\)](#) holds, then there exists  $\alpha > 0$  small enough such that*

(a) *the optimal labor force is majority-dominant but over-represents the minority relative to its population share,  $\phi^* \in (0.5, b)$ .*

(b) *a sufficiently patient planner intervenes persistently in favor of the minority.*

*Proof.* See [Appendix A](#). □

Coincidentally, situations with small  $\alpha$  are exactly the ones that [Athey et al. \(2000, Prop. 2 and 4\)](#) identify as the settings where a fully balanced labor force is both optimal and a stable steady state without intervention. In other words: Extrapolating from a model with equal population pools 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 situation where surplus maximization requires ongoing intervention.

To illustrate this robustness with an example, [Figure 6](#) assumes normally distributed talent (panel a) and a different mentoring technology (panel b). As before, match opportunities arise randomly with probability  $\frac{q}{nl}$ , and a junior receives

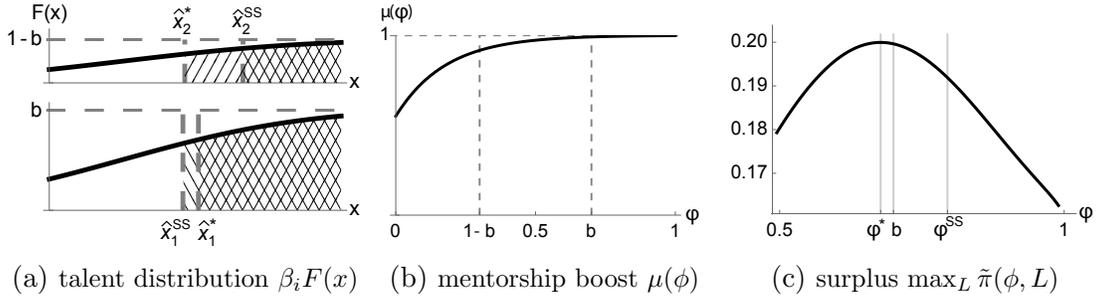


Figure 6: Over-representation of the minority under normally distributed talent with cross-group mentoring benefits.

a mentorship boost of 1 if at least one of the matches is successful. However, we no longer assume that opportunities lead to a successful match if and only if both sides belong to the same group. Instead, we assume that a possible match is successful with probability  $p^H = 0.8$  if agents belong to the same group, and  $p^L = 0.1$  if they do not.<sup>32</sup> Panel c reveals that the optimal composition  $\phi^*$  over-represents the minority and improves total surplus by 4.20% relative to the economy's unique steady state  $\phi^{SS}$ .<sup>33</sup>

## 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. However, we believe that awareness of the dynamic consequences of mentoring complementarities is crucial for the public discussion. On the most basic level, the traits 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 skills, it makes sense to remunerate group-specific mentoring ability according to the shortness of

<sup>32</sup>As  $n \rightarrow \infty$ , the expected mentorship boost at a steady state is then given by  $\mu(\phi) = 1 - e^{-q(\phi p^H + (1-\phi)p^L)}$ .

<sup>33</sup>We assume  $c = 4.4$ ,  $\alpha = 3$ ,  $\beta_1 = b = 0.7$ ,  $\beta_2 = 1 - b = 0.3$ , and  $q = 8$ . Note that  $\alpha$  is far from marginal, and mentoring holds significant importance in this example. Together with  $\bar{x} = -5$ ,  $\bar{x} = 1.7$  and  $k = 0.14$ , this satisfies all conditions of Lemma 6 and hence the optimal labor force over-represents the minority.

its supply and its impact on future surplus. However, such remuneration does not arise in an unregulated economy due to firm competition, and minority workers fail to 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. In sectors that require rare skills, the forward-looking optimal remuneration 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.

Natural follow-up questions remain. First, we do not demonstrate or quantify the strength of mentoring externalities. Their existence has been documented empirically by studies mentioned in the introduction, but more research is needed to get reliable estimates of their salience. Second, we assume a fixed ‘mentoring technology’ and take these cross-racial or cross-gender mentoring hurdles as given. We do not ask ‘What if mentoring *itself* could be improved?’ because we believe that estimating the cost of such improvements is mainly an empirical question. It is important to highlight that programs which facilitate mentor assignments for minority juniors, or improve cross-group mentoring skills, tend to decrease access hurdles in a way similar to minority scholarships. Policy makers have been fostering such mentoring for minority youth,<sup>34</sup> and the most successful diversity programs are exactly those that increase cross-group exposure.<sup>35</sup> These approaches

---

<sup>34</sup>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>).

<sup>35</sup>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.

chip away at the same obstacles that we study, albeit from an angle that is not considered here.

## A Additional Proofs

**Proof of Proposition 1.** Consider a total student body of size  $l$ , and let

$$\ell_i(l) = \begin{cases} 0 & \text{if } c - \lambda - \alpha\mu(L_i/l) > w_i \\ \infty & \text{if } c - \alpha\mu(L_i/l) < w_i \\ \beta_i x^{-1}(c - w_i - \alpha\mu(L_i/l)) & \text{otherwise.} \end{cases}$$

The equilibrium conditions (3) are equivalent to  $l = \ell_1(l) + \ell_2(l)$ . Mentoring  $\mu(L_i(t)/l)$  is strictly decreasing in  $l$ , and therefore so is  $\ell_1(l) + \ell_2(l)$  over the range  $(0, \infty)$ . In other words, as  $l$  grows, school enrollment  $\ell_1(l) + \ell_2(l)$  weakly drops, implying a single crossing  $l = \ell_1(l) + \ell_2(l)$ .

Similarly,  $\ell_i(l)$  is increasing in  $w_i$  and independent of  $w_j$ . A raise in  $w_i$  thus shifts total school enrollment  $\ell_1(l) + \ell_2(l)$  upwards, and the single crossing to the right. The extra supply comes entirely from group  $i$ ; group- $j$  enrollment  $\ell_j(l)$  drops due to the increase in  $l$ .  $\square$

**Lemma 2.** *At a homogeneous steady state with group  $i$  working, total labor supply is given by  $L = \beta_i x^{-1}(c - 1 - \alpha\mu(1))$ . At a mixed steady state, total labor supply is  $L = \frac{\beta_1 + \beta_2}{\lambda} (\ln(\lambda) - bg(\phi) - (1 - b)g(1 - \phi))$ , where  $g(\phi) = \ln(c - 1 - \alpha\mu(\phi))$ . If  $\phi \neq b$ , this is equal to  $\frac{b(1-b)}{\lambda(b-\phi)}(g(\phi) - g(1 - \phi))(\beta_1 + \beta_2)$ .*

*Proof.* At a homogeneous steady state, labor force participation is obtained directly by inverting the relevant equation in (3). At a mixed steady state, total participation is given by  $L = \beta_1 x^{-1}(c - 1 - \alpha\mu(\phi)) + \beta_2 x^{-1}(c - 1 - \alpha\mu(1 - \phi))$ , which simplifies to the expression above.  $\square$

**Lemma 3.** *The function  $g(\phi) = \ln(c - 1 - \alpha\mu(\phi))$  is decreasing and convex under Assumption (A1<sup>+</sup>). As  $q \rightarrow \infty$ ,  $g(\phi) \rightarrow \ln(c - 1 - \alpha)$  and  $g'(\phi) \rightarrow 0$  pointwise for all  $\phi \in (0, 1)$ . Finally,  $\partial g(\phi)/\partial q = \frac{\phi}{q}g'(\phi)$  is increasing in  $\phi$ .*

*Proof.* Negative monotonicity follows immediately from monotonicity of  $\mu$  and the logarithm function. Convexity follows since

$$g''(\phi) = \frac{\alpha q e^{-q\phi}}{c - 1 - \alpha\mu(\phi)} q \left( 1 - \frac{\alpha(1 - \mu(\phi))}{c - 1 - \alpha\mu(\phi)} \right) > 0$$

by [Assumption \(A1<sup>+</sup>\)](#).

The limits are a direct consequence of  $\lim_{q \rightarrow \infty} \mu(\phi) = 1$  and  $\lim_{q \rightarrow \infty} \mu'(\phi) = 0$ .

Finally,  $\partial\mu(\phi)/\partial q = \phi e^{-q\phi} = \frac{\phi}{q} \mu'(\phi)$  implies the expression for the partial derivative. It is increasing in  $\phi$  by convexity of  $g$ .  $\square$

**Supplement to the proof of [Lemma 1](#).** Consider a steady state with composition  $\phi \in (0, 1)$ . By the argument in the main text,  $S(\phi) = 0$ . It remains to show that  $\phi$  is stable whenever  $S'(\phi) < 0$ .

The equation of motion  $\dot{\mathbf{L}} = \mathbf{l}(\mathbf{L}) - \mathbf{L}$  is  $C^1$  over  $\mathbb{R}_+^2$ . By the Linearization Theorem, the steady state is therefore stable if and only if all eigenvalues of the Jacobian  $\frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}}$  have a negative real part. For simplicity, we first rewrite the equations in [\(3\)](#) as

$$F(\mathbf{l}, \mathbf{L}) = \begin{bmatrix} b \left( \ln \lambda - g \left( \frac{L_1}{l_1 + l_2} \right) \right) - \frac{\lambda l_1}{\beta_1 + \beta_2} \\ (1 - b) \left( \ln \lambda - g \left( \frac{L_2}{l_1 + l_2} \right) \right) - \frac{\lambda l_2}{\beta_1 + \beta_2} \end{bmatrix} = \mathbf{0},$$

where  $g(\phi)$  is as in [Lemma 3](#).

The Implicit Function Theorem implies that at the steady state  $(\phi L, (1 - \phi)L)$ ,

$$\begin{aligned} \frac{\partial \mathbf{l}}{\partial \mathbf{L}} &= - \left[ \frac{\partial F}{\partial \mathbf{l}} \right]^{-1} \frac{\partial F}{\partial \mathbf{L}} \\ &= \begin{bmatrix} \frac{b\phi}{1-b} g'(\phi) - \frac{\lambda L}{\beta_2} & \frac{b\phi}{1-b} g'(\phi) \\ (1-\phi)g'(1-\phi) & (1-\phi)g'(1-\phi) - \frac{\lambda L}{\beta_2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{b}{1-b} g'(\phi) & 0 \\ 0 & g'(1-\phi) \end{bmatrix}. \end{aligned}$$

After simplification using [Lemma 2](#), the characteristic polynomial  $\left| \frac{\partial \dot{\mathbf{L}}}{\partial \mathbf{L}} - \gamma I \right|$  is proportional to  $F(\gamma) = A\gamma^2 + B\gamma + C$  with

$$A = \frac{\lambda L}{\beta_2} \left( -(1-b) \frac{\lambda L}{\beta_2} + (1-b)(1-\phi)g'(1-\phi) + b\phi g'(\phi) \right) < 0$$

$$\begin{aligned}
B &= -\frac{2}{1-b} \left[ S'(\phi) + \frac{S(\phi)}{b-\phi} \right]^2 \\
&\quad - \left[ S'(\phi) + \frac{S(\phi)}{b-\phi} \right] \left[ (1+2\phi)g'(1-\phi) + (3-2\phi)\frac{b}{1-b}g'(\phi) \right] \\
&\quad - \left[ (1-b)\phi g'(1-\phi)^2 + \frac{b^2}{1-b}(1-\phi)g'(\phi)^2 \right] \\
C &= \frac{\lambda L}{\beta_2} \left[ S'(\phi) + \frac{S(\phi)}{b-\phi} \right].
\end{aligned}$$

This is a downward sloping quadratic function since  $A < 0$ . The real part of its roots are negative if and only if  $F(0) = C < 0$  and  $F'(0) = B < 0$ . At a steady state,  $S(\phi) = 0$  and hence  $C < 0$  if and only if  $S'(\phi) < 0$ , which in turn implies  $B < 0$ .  $\square$

**Proof of Theorem 1.** The first part of the theorem is immediate: When costs are so low that the most able individuals invest even without *any* mentorship, their labor supply never dries out. Under [Property \(hSS\)](#) however, no workers get educated unless mentor availability exceeds some positive threshold  $\underline{\Lambda}$ . This ensures that the opposing group has a steady-state mentorship boost of  $\mu(1)$ , which determines labor supply through [Equation \(3\)](#). Such a homogeneous steady state is stable since small enough perturbations maintain minority mentor availability below  $\underline{\Lambda}$ .

Let us now turn to mixed steady states  $\mathbf{l} = \mathbf{L} = (\hat{\phi}\hat{L}, (1-\hat{\phi})\hat{L}) \in \mathbb{R}_+^2$ , where the individual cost-benefit analyses in [Equation \(3\)](#) simplify to

$$c - x \left( \frac{\hat{\phi}\hat{L}}{\beta_1} \right) - \alpha\mu(\hat{\phi}) = 1 \quad \text{and} \quad c - x \left( \frac{(1-\hat{\phi})\hat{L}}{\beta_2} \right) - \alpha\mu(1-\hat{\phi}) = 1. \quad (11)$$

The proof makes heavy use of the auxiliary function  $S$  defined in [Lemma 1](#), whose roots pin down steady state composition. It is easily verified that [Property \(mSS\)](#) is *necessary* for a steady supply of minority workers, for otherwise either at least one of the left-side expressions exceeds 1. As for sufficiency, note that  $S$  is continuous and  $(1-b)S(0) = -bS(1)$ . For  $S(1) \neq 0$ , or, equivalently,  $\lambda \neq c-1$ , the change of

sign implies that  $S$  admits an interior root.<sup>36</sup>

The limiting behavior of  $g$  implies the existence of a unique stable mixed steady state near  $b$  for  $q$  or  $\lambda$  large enough. We first show existence, and start with the case  $q \rightarrow \infty$  first. It follows from [Lemma 3](#) that

$$S(b) = (1 - b)b(g(1 - b) - g(b)) > 0 \quad (12)$$

vanishes as  $q$  grows, and its derivative admits a negative limit

$$\begin{aligned} S'(b) &= (1 - b)g(1 - b) + bg(b) - \ln \lambda - (1 - b)b(g'(1 - b) + g'(b)) \\ &\xrightarrow{q \rightarrow \infty} k = \ln(c - 1 - \alpha) - \ln(\lambda) \stackrel{(A1)}{<} 0. \end{aligned}$$

The first order Taylor approximation implies that for  $\delta > 0$  arbitrarily small,

$$S(b + \delta) \leq S(b) + \delta S'(b) - \frac{k}{2}\delta.$$

The limit  $\lim_{q \rightarrow \infty} S(b) + \delta S'(b) = \delta k$  implies that there exists  $Q_1$  large enough such that

$$S(b + \delta) \leq S(b) + \delta S'(b) - \frac{k}{2}\delta < \frac{k}{2}\delta - \frac{k}{2}\delta = 0 \quad \forall q \geq Q_1. \quad (13)$$

More generally,  $S(b + \delta) = -\delta \ln \lambda + (1 - b - \delta)bg(b + \delta) + (b + \delta)(1 - b)g(1 - b - \delta)$  is decreasing in  $\lambda$  and unbounded below. Hence, there exists  $\Lambda_1 > 0$  such that

$$S(b + \delta) < 0 \quad \forall \lambda > \Lambda_1. \quad (13')$$

By continuity, [Equation \(12\)](#) and either [\(13\)](#) or [\(13'\)](#) imply that  $S(\phi)$  crosses 0 downwards at some  $\phi \in (b, b + \delta)$ . By [Lemma 1](#), this crossing constitutes a stable steady state.

As for uniqueness, note first that  $\varepsilon = \min \left\{ 1, \frac{1}{\alpha} \sqrt{c - 1 - \alpha} \left( \sqrt{\lambda} - \sqrt{c - 1 - \alpha} \right) \right\}$  is positive by [Assumptions \(A1\)](#) and [\(A1<sup>+</sup>\)](#). For large  $q$  or  $\lambda$ , we now partition

---

<sup>36</sup>In the special case that  $\lambda = c - 1$ , it follows that  $S'(1) = b(g(1) - g(0)) - (1 - b)g'(0) > g'(0)(2b - 1) \geq 0$ , where the inequalities hold by convexity and negative monotonicity of  $g(\phi)$  (see [Lemma 3](#)). Together with  $S(b) = (1 - b)b(g(1 - b) - g(b)) \geq 0$ , this also implies a zero over  $[b, 1)$ .

$[0, 1]$  into five intervals and pin down the location stable steady state in one of them.

First, consider the middle range  $I_3 = [\frac{1-b}{2}, \frac{1+b}{2}]$  and note that

$$\begin{aligned} S'(\phi) &= -\ln(\lambda) + b \ln(c-1-\alpha\mu(\phi)) + (1-\phi)b \frac{\alpha\mu'(\phi)}{c-1-\alpha\mu(\phi)} \\ &\quad + (1-b) \ln(c-1-\alpha\mu(1-\phi)) + \phi(1-b) \frac{\alpha\mu'(1-\phi)}{c-1-\alpha\mu(1-\phi)} \\ &< -\ln(\lambda) + \ln\left(c-1-\alpha\mu\left(\frac{1-b}{2}\right)\right) + \frac{\alpha\mu'\left(\frac{1-b}{2}\right)}{c-1-\alpha\mu\left(\frac{1-b}{2}\right)}. \end{aligned}$$

This is unbounded below in  $\lambda$  and converges to  $-\ln(\lambda) + \ln(c-1-\alpha) < 0$  as  $q \rightarrow \infty$ . Consequently, there exist  $\Lambda_2$  and  $Q_2$  such that  $S$  is strictly decreasing over  $I_3$  for either  $\lambda > \Lambda_2$  or  $q > Q_2$ . A strictly decreasing function can have at most one root.

Outside of this range, for  $\lambda > 1$ ,  $|S(\phi)|$  is uniformly bounded below,

$$\begin{aligned} |S(\phi)| &= |(b-\phi)\ln(\lambda) - (1-\phi)b\ln(c-1-\alpha\mu(\phi)) + \phi(1-b)\ln(c-1-\alpha\mu(1-\phi))| \\ &> |b-\phi|\ln(\lambda) - \ln(c-1) > \frac{1-b}{2}\ln(\lambda) - \ln(c-1) \xrightarrow{\lambda \rightarrow \infty} -\infty. \end{aligned}$$

As a consequence, there exists  $\Lambda_3 > 0$  such that  $S$  admits no further roots for all  $\lambda > \Lambda_3$ . Alternatively, consider any  $q > Q_3 = \frac{c-1}{\alpha\varepsilon(1-b)} \ln\left(\frac{\lambda}{c-1-\alpha}\right) - \ln\varepsilon > 0$ . Let  $\phi_0 = \mu^{-1}(1-\varepsilon) = -\frac{1}{q}\ln\varepsilon$ , and note that  $\mu'(\phi_0) = q\varepsilon$ . Over the intervals  $I_1 \cup I_5 = ([0, \phi_0] \cup [1-\phi_0, 1]) \cap [0, 1]$ ,

$$\begin{aligned} S'(\phi) &> -\ln(\lambda) + \ln(c-1-\alpha) + (1-\phi_0)(1-b) \frac{\alpha\mu'(\phi_0)}{c-1} \\ &= -\ln(\lambda) + \ln(c-1-\alpha) + (1-b) \frac{\alpha Q_3 \varepsilon}{c-1} + \ln\varepsilon(1-b) \frac{\alpha\varepsilon}{c-1} = 0. \end{aligned}$$

In other words, any steady states over this range are unstable.

Finally, by the definition of  $\varepsilon$ ,  $S(\phi)$  is positive over  $I_2 = [\phi_0, \frac{1-b}{2}]$ ,

$$\begin{aligned} S(\phi) &= (b-\phi)\ln(\lambda) - (1-\phi)b\ln(c-1-\alpha\mu(\phi)) + \phi(1-b)\ln(c-1-\alpha\mu(1-\phi)) \\ &> (b-\phi)\ln(\lambda) - (1-\phi)b\ln(c-1-\alpha\mu(\phi_0)) + \phi(1-b)\ln(c-1-\alpha) \end{aligned}$$

$$> \frac{b}{2} \ln \left( \frac{\lambda}{c-1-\alpha} \right) - b \ln \left( 1 + \frac{\alpha\varepsilon}{c-1-\alpha} \right) = 0$$

and negative over  $I_4 = [\frac{1+b}{2}, 1 - \phi_0]$ ,

$$\begin{aligned} S(\phi) &= (b - \phi) \ln(\lambda) - (1 - \phi)b \ln(c - 1 - \alpha\mu(\phi)) + \phi(1 - b) \ln(c - 1 - \alpha\mu(1 - \phi)) \\ &< (b - \phi) \ln(\lambda) - (1 - \phi)b \ln(c - 1 - \alpha) + \phi(1 - b) \ln(c - 1 - \alpha\mu(\phi_0)) \\ &< -\frac{1-b}{2} \ln \left( \frac{\lambda}{c-1-\alpha} \right) + (1-b) \ln \left( 1 + \frac{\alpha\varepsilon}{c-1-\alpha} \right) = 0. \end{aligned}$$

For either  $q > \max\{Q_1, Q_2, Q_3\}$  or  $\lambda > \max\{\Lambda_1, \Lambda_2\}$ , there exists therefore a unique stable steady state  $\phi$ , with  $\phi \in (b, b + \delta)$ .

Over-representation of the majority can be shown for  $b > 0.5$  in two steps based on [Lemma 1](#). First,  $S$  admits no root over  $\phi \in [0.5, b]$  since

$$S(\phi) = (b-\phi)x^{-1}(c-1-\alpha\mu(\phi)) + \phi(1-b)(x^{-1}(c-1-\alpha\mu(\phi)) - x^{-1}(c-1-\alpha\mu(1-\phi))) > 0.$$

Indeed, by [Property \(mSS\)](#) and monotonicity of  $x$ , both terms are nonnegative. The first term is strictly positive for  $\phi < b$ , the second for  $\phi > 0.5$ . Second, we establish positivity of the following expression of  $g(\phi) = \ln(c - 1 - \alpha\mu(\phi))$  and  $h(\phi) = g(1 - \phi) - g(\phi)$  whenever  $0 < \phi < 0.5 < b$ ,

$$\begin{aligned} (b - \phi)S'(\phi) + S(\phi) &= (1 - b)b \cdot h(\phi) - (b - \phi)((1 - b)\phi g'(1 - \phi) + b(1 - \phi)g'(\phi)) \\ &\geq \frac{1}{4}h(\phi) - \left(\frac{1}{2} - \phi\right) \left(\frac{\phi}{2}g'(1 - \phi) + \frac{1 - \phi}{2}g'(\phi)\right) \\ &\geq \frac{1}{4} \left[ h(\phi) + \left(\frac{1}{2} - \phi\right) h'(\phi) \right] \stackrel{(mSS^+)}{>} \frac{1}{4}h\left(\frac{1}{2}\right) = 0. \end{aligned}$$

The first inequality holds because the expression is decreasing in  $b$ . The second inequality owes to convexity of  $g$ , and hence  $g'(1 - \phi) > g'(\phi)$ . Finally, [Property \(mSS<sup>+</sup>\)](#) ensures concavity of  $h(\phi)$  over  $(0, 0.5)$  and implies the last inequality. Together, these inequalities imply that any steady state ( $S(\phi) = 0$ ) with group-1 in the minority ( $b - \phi > 0.5 - \phi > 0$ ) must have positive slope  $S'(\phi) > 0$ , and hence be unstable.  $\square$

**Proof of Theorem 2.**  $S$  admits the following partial derivatives:

$$\begin{aligned}\frac{\partial S}{\partial b} &= (1 - \phi)(\ln \lambda - g(\phi)) + \phi(\ln \lambda - g(1 - \phi)) > 0, \\ \frac{\partial S}{\partial \lambda} &= (b - \phi)/\lambda < 0, \\ \frac{\partial S}{\partial q} &= (1 - \phi)b \frac{\partial g(\phi)}{\partial q} + \phi(1 - b) \frac{\partial g(1 - \phi)}{\partial q} \\ &= b \frac{(1 - \phi)\phi}{q} g'(\phi) + (1 - b) \frac{(1 - \phi)\phi}{q} g'(1 - \phi) < 0.\end{aligned}$$

which owes to the properties of  $g(\phi)$  (see Lemma 3) and the positive labor supply in any mixed steady state, which implies  $\ln \lambda > g(1 - \phi) > g(\phi)$ . The sign of these derivatives implies that  $S(\phi) < 0$  for any of the changes mentioned in the statement.

In addition,  $S$  remains positive at  $b > 0.5$  since  $S(b) = (1 - b)b(g(1 - b) - g(b)) > 0$  by Lemma 3. Continuity of  $S$  then implies the existence of a downward-crossing (and by Lemma 1 a stable steady state) over  $(b, \phi) \subseteq (0.5, \phi)$ .  $\square$

**Proof of Theorem 3.** Define the single-group surplus

$$\pi_i(m) = \tilde{\pi}(2 - i, \beta_i x^{-1}(c - 1 - \alpha m)) = \beta_i \left[ 1 - \left( 1 - \ln \left( \frac{c - 1 - \alpha m}{\lambda} \right) \right) \frac{c - 1 - \alpha m}{\lambda} \right]$$

for any mentoring level that attracts at least some students,  $c - 1 - \alpha m < \lambda$ . Note that this quantity is smaller than  $\beta_i$  and increasing in  $m$ . It is strictly positive since  $1 - \ln(y) < \frac{1}{y}$  for  $y = \frac{c - 1 - \alpha m}{\lambda} \in (0, 1)$  by concavity of the logarithm.<sup>37</sup> As  $\lambda \rightarrow \infty$ , most terms disappear and  $\pi_i(m) \rightarrow \beta_i$  pointwise for all values of  $m$ . Similarly, as  $q \rightarrow \infty$ ,  $\mu(\phi) \rightarrow 1$  and hence  $\pi_i(\mu(\phi)) \rightarrow \pi_i(1)$  pointwise for all values of  $\phi$ .

In Theorem 1, we establish the existence of a stable steady state arbitrarily close to composition  $b$  for large enough  $q$  or  $\lambda$ . The surplus of that steady state

---

<sup>37</sup>The first-order Taylor approximation yields  $0 = \ln(1) < \ln(y) + (1 - y) \ln'(y) = \ln(y) + \frac{1}{y} - 1$ .

eventually exceeds that of a homogeneous workforce since

$$\pi_1(\mu(b)) + \pi_2(\mu(1-b)) \rightarrow \beta_1 + \beta_2 > \beta_i > \pi_i(\mu(1)) \quad \text{as } \lambda \rightarrow \infty$$

and

$$\pi_1(\mu(b)) + \pi_2(\mu(1-b)) \rightarrow \pi_1(1) + \pi_2(1) > \pi_i(\mu(1)) \quad \text{as } q \rightarrow \infty.$$

Eventually, this stable mixed steady state therefore dominates the homogeneous ones in terms of social surplus. There may be an even better stable steady state, but it must be mixed.  $\square$

The next lemma establishes that the optimal long-term total labor force doesn't explode or disappear. Under [Assumption \(A1<sup>+</sup>\)](#), it establishes bounds on the  $L^*$  that hold for all valid parameter values, i.e. all those  $q$  that satisfy [Assumption \(A1\)](#).

**Lemma 4.** *Long-term surplus  $\tilde{\pi}(\phi, L)$  attains its maximum  $(\phi^*, L^*)$  at some  $\phi \in [0.5, 1]$  and  $L^* \in (0, \bar{L}]$  for  $\bar{L} = \frac{\beta_1 + \beta_2}{c - 1 - \alpha\mu(1)}$ . Under [Assumption \(A1<sup>+</sup>\)](#) and for  $q > Q > \ln(\alpha) - \ln(\lambda - c + 1 + \alpha)$ , the optimal labor force is contained in  $[\underline{L}, \frac{\beta_1 + \beta_2}{c - 1 - \alpha}]$ , where  $\underline{L} > 0$  only depends on  $Q$ .*

*Proof.* The function  $\rho(L_i, \beta_i, L) = \beta_i(1 - e^{-\lambda \frac{L_i}{\beta_i}}) + L_i(1 - c + \alpha\mu(L_i/L))$  is strictly supermodular in its first two arguments since  $\frac{\partial^2 \rho}{\partial L_i \partial \beta_i} = \lambda^2 / \beta_i^2 L_i e^{-\lambda \frac{L_i}{\beta_i}} > 0$ . Since  $\tilde{\pi}(\phi, L) = \rho(\phi L, \beta_1, L) + \rho((1 - \phi)L, \beta_2, L)$ , this implies that the optimal labor force leans towards the majority group,  $\phi^* \geq 0.5$ .

Note that  $\bar{L}$  is positive by [Assumption \(A1\)](#) and increasing in  $q$ . By continuity and the Extreme Value Theorem,  $\tilde{\pi}$  attains a supremum  $\tilde{\pi}(\phi^*, L^*)$  over the compact domain  $[0, 1] \times (0, \bar{L}]$ . This supremum is strictly positive and  $L^* > 0$  since  $\tilde{\pi}(\phi, 0) \equiv 0$  and  $\frac{\partial}{\partial L} \tilde{\pi}(1, L) = \lambda + 1 - c + \alpha\mu(1) > 0$  by [Assumption \(A1\)](#). It is a global maximum since any larger labor force  $L > \bar{L}$  leads to negative surplus

$$\tilde{\pi}(\phi, L) < \beta_1 + \beta_2 + \underbrace{\phi L (1 - c + \alpha\mu(\phi))}_{< 1 - c + \alpha\mu(1) < 0} + (1 - \phi)L \underbrace{(1 - c + \alpha\mu(1 - \phi))}_{< 1 - c + \alpha\mu(1) < 0}$$

$$< \beta_1 + \beta_2 + \tilde{L}(1 - c + \alpha\mu(1)) < 0.$$

For the lower bound, choose any  $Q > \ln(\alpha) - \ln(\lambda - c + 1 + \alpha)$  arbitrary. The choice of  $Q$  ensures that [Assumption \(A1\)](#) holds. For a single working group, long-term surplus under  $Q$  reduces to  $\tilde{\pi}(1, L; Q) = \int_0^L 1 - c + x(l/\beta_1) + \alpha\mu(1; Q)dl$ . Its integrand is strictly decreasing in  $l$ , and zero exactly when  $l = L_1^* = \beta_1 x^{-1}(c - 1 - \alpha\mu(1; Q))$ . Surplus is increasing in  $q$  and hence any optimal long-term surplus is bounded below by  $\tilde{\pi}(1, L_1^*; Q) > 0$ . Surplus is also bounded above by  $L \cdot (1 - c + \lambda + \alpha)$  since the second term bounds each integrand. Together, these imply that  $\tilde{\pi}(1, L_1^*; Q) \leq \tilde{\pi}(1, L_1^*; q) \leq \tilde{\pi}(\phi^*, L^*; q) \leq L^*(1 - c + \lambda + \alpha)$  or, equivalently,  $L^* > \underline{L} = \frac{\tilde{\pi}(1, L_1^*; Q)}{1 - c + \lambda + \alpha}$  for all  $q > Q$ .  $\square$

**Proof of [Theorem 4](#).** The existence of a surplus-maximizing labor force with positive supply and group-1 majority follows directly from [Lemma 4](#). As for the dependence on mentor capacity, we prove each claim in turn:

- (a) Formally, we show that there exists  $q_0 > \underline{q} > 0$  such that [Assumption \(A1\)](#) holds for all  $q > \underline{q}$  and  $\phi^* = 1$  for all  $q \in [\underline{q}, q_0]$ . To do so, let  $\underline{q}$  and  $q_0$  be the unique solutions to  $c - 1 - \lambda - \alpha\mu(1) = 0$  and  $c - 1 - \lambda - \alpha\mu(0.5) = 0$  respectively. These solutions exist and are unique because the left-side expressions are strictly decreasing in  $q$ , they tend to  $c - 1 - \lambda > 0$  as  $q \rightarrow 0$  by [Property \(hSS\)](#) and to  $c - 1 - \lambda - \alpha < 0$  as  $q \rightarrow \infty$  whenever [\(A1\)](#) holds for any  $q$ . They are ordered as  $q_0 > \underline{q}$  since  $\mu$  is strictly increasing in  $\phi$  for any positive  $q$ .

For  $q \in [\underline{q}, q_0]$ , minority participation is always inefficient. To see this, consider any  $\phi \in (0.5, 1)$  and note that  $1 - c + \lambda + \alpha\mu(1 - \phi) < 0$  since  $q < q_0$ . In other words, the marginal surplus from even the most able minority worker is negative. Excluding them also improves the mentorship boost for the majority, and hence

$$\tilde{\pi}(\phi, L) = \underbrace{\rho(\phi L, \beta_1, L)}_{< \rho(\phi L, \beta_1, \phi L)} + \underbrace{\rho((1 - \phi)L, \beta_2, L)}_{< 0 = \rho(0, \beta_2, \phi L)} < \tilde{\pi}(1, \phi L),$$

where  $\rho$  is as in Lemma 4. This implies that the optimal labor force is homogeneous.

- (b) The function  $M(\phi) = \phi e^{-q\phi} + (1 - \phi)e^{-q(1-\phi)}$  is the weighted sum of two convex exponential functions. As such, it is bounded below by taking a first-order Taylor approximation of each term at  $\phi = \phi_0$ ,

$$\begin{aligned} M(\phi) &\geq \phi(e^{-q\phi_0} + (\phi - \phi_0)(-qe^{-q\phi_0})) + (1 - \phi)(e^{-q(1-\phi_0)} + (\phi - \phi_0)(qe^{-q(1-\phi_0)})) \\ &= e^{-q\phi_0}\phi(1 - q(\phi - \phi_0)) + e^{-q(1-\phi_0)}(1 - \phi)(1 + q(\phi - \phi_0)) = \underline{M}(\phi, \phi_0). \end{aligned}$$

Similarly,  $N(\phi, L) = \beta_1 e^{-\lambda \frac{\phi L}{\beta_1}} + \beta_2 e^{-\lambda \frac{(1-\phi)L}{\beta_2}}$  is the sum of two convex exponential functions. It is bounded below by the second-order Taylor approximation at  $\phi = \phi_0$ ,

$$\begin{aligned} N(\phi, L) &\geq e^{-\lambda \frac{\phi_0 L}{\beta_1}} \left( \beta_1 - (\phi - \phi_0)\lambda L + (\phi - \phi_0)^2 \frac{(\lambda L)^2}{\beta_1} \right) \\ &\quad + e^{-\lambda \frac{\phi_0 L}{\beta_2}} \left( \beta_2 + (\phi - \phi_0)\lambda L + (\phi - \phi_0)^2 \frac{(\lambda L)^2}{\beta_2} \right) = \underline{N}(\phi, \phi_0, L). \end{aligned}$$

Long-term surplus can be written as

$$\tilde{\pi}(\phi, L) = \beta_1 + \beta_2 + L(1 - c + \alpha) - N(\phi, L) - \alpha LM(\phi).$$

Replacing  $M$  and  $N$  with  $\underline{M}$  and  $\underline{N}$  yields a quadratic upper bound  $\bar{\pi}$ , for any  $\phi_0$ .

Let now  $\phi_0 = b$ . For  $q > Q_1$  large enough, the coefficient of the leading term  $\phi^2$ ,

$$\left. \frac{\partial^2 \bar{\pi}}{\partial \phi^2} \right|_{\phi_0=b} = \underbrace{-(\lambda L)^2 \left( \frac{1}{\beta_1} + \frac{1}{\beta_2} \right) e^{-\frac{\lambda L}{\beta_1 + \beta_2}}}_{\text{negative and independent of } q} + \underbrace{\alpha L q (e^{-qb} + e^{-q(1-b)})}_{\rightarrow 0 \text{ as } q \rightarrow \infty}, \quad (14)$$

is negative, making the upper bound concave. The first derivative

$$\begin{aligned}\frac{\partial \tilde{\pi}}{\partial \phi} \Big|_{\phi_0=b} &= \frac{\partial \bar{\pi}}{\partial \phi} \Big|_{\phi_0=b} = - \underbrace{N_1(b, b, L)}_{=0} - \alpha L \underline{M}_1(b, b) \\ &= -\alpha L (e^{-qb}(1 - qb) - e^{-q(1-b)}(1 - q(1 - b))) < 0\end{aligned}$$

is also negative for  $q > Q_2$  large enough.<sup>38</sup>

We conclude as follows: For  $q > \max\{Q_1, Q_2\}$  large enough, surplus is locally decreasing at  $\phi = b$  since  $\frac{\partial \tilde{\pi}}{\partial \phi}(\phi, L) < 0$ . Together with the concave upper bound that is tangent at  $\phi = b$ , this implies that there exists a small enough  $\delta > 0$  such that

$$\tilde{\pi}(b - \delta, L) > \tilde{\pi}(b, L) = \bar{\pi}(b, L) > \bar{\pi}(\phi, L) \geq \tilde{\pi}(\phi, L) \quad \forall \phi \in [b, 1].$$

In particular, this holds at the optimal total labor participation  $L^*$ .

- (c) To show convergence of the optimal labor force composition, fix any  $\varepsilon > 0$ . By [Lemma 4](#) and [Assumption \(A1<sup>+</sup>\)](#), the optimal total labor force admits an upper and lower bound that is independent of  $q$ ,  $\underline{L} < L^* < \bar{L}$  for all  $q > Q_3 > \ln(\alpha) - \ln(\lambda - c + 1 + \alpha)$  large enough. We have already established that  $\frac{\partial N}{\partial \phi}(b, L) \equiv 0$  for any  $L > 0$ . Moreover, by [Equation \(14\)](#) and continuity of  $\frac{\beta_1 + \beta_2}{b(2-b)}(\lambda L)^2 e^{-\lambda L}$ , there exists  $\kappa > 0$  and  $Q_4 > 0$  such that

$$\frac{\partial^2 \bar{\pi}}{\partial \phi^2}(\phi, L^*) < -\kappa \quad \forall \phi, \forall L \in [\underline{L}, \bar{L}] \text{ and } \forall q > Q_4.$$

Also, by convexity of the quadratic function  $\underline{M}$ , its derivative converges uniformly to zero since  $\underline{M}'(1) \leq \underline{M}'(\phi) \leq \underline{M}'(0)$  and  $\lim_{q \rightarrow \infty} (c_1 + c_2 q) e^{-c_3 q}$  for any  $c_1, c_2 \in \mathbb{R}$  and  $c_3 > 0$ . As a consequence, there exists  $Q_5 > 0$  such that  $|\underline{M}'(\phi)| < \frac{\kappa \varepsilon}{\alpha} \bar{L}^{-1}$  for all  $q > Q_5$ .

Whenever  $|b - \phi| > \varepsilon$  and  $q > \max\{Q_3, Q_4, Q_5\}$ , let  $L^* = \arg \max_L \{\tilde{\pi}(\phi, L)\}$ .

---

<sup>38</sup>For  $q$  large enough, the exponential terms in  $M'(b) = e^{-qb}(1 - qb - e^{q(2b-1)}(1 - q(1 - b)))$  dominate, eventually rendering both factors positive.

It follows that

$$\tilde{\pi}(\phi, L^*) \leq \bar{\pi}(\phi, L^*) = \bar{\pi}(b, L^*) + (\phi - b) \frac{\partial \bar{\pi}}{\partial \phi}(b, L^*) + (\phi - b)^2 \frac{\partial^2 \bar{\pi}}{\partial \phi^2}(b, L^*)$$

for the quadratic upper envelope  $\bar{\pi}$ . At  $\phi = b$ , the two are tangent and the first derivative simplifies, hence

$$\bar{\pi}(\phi, L^*) = \tilde{\pi}(b, L^*) - \alpha(\phi - b)L^*M'(\phi) + (\phi - b)^2 \frac{\partial^2 \bar{\pi}}{\partial \phi^2}(b, L^*)$$

Finally, the bounds on  $M'$  and the concavity of  $\bar{\pi}$  imply

$$\bar{\pi}(\phi, L^*) < \tilde{\pi}(b, L^*) + \kappa\varepsilon|\phi - b| - (\phi - b)^2\kappa < \tilde{\pi}(b, L^*).$$

In other words, composition  $\phi$  is strictly dominated by  $b$ , and  $\lim_{q \rightarrow \infty} \phi^* = b$ .

- (d) Let now  $\phi_0 = \hat{\phi}$ , and use a similar approach as in part (b). Since  $\hat{\phi}$  is a mixed steady state, [Equation \(3\)](#) implies for the leading coefficient

$$\begin{aligned} \frac{\partial^2 \bar{\pi}}{\partial \phi^2} &= -\frac{\lambda L^2}{\beta_1} x \left( \frac{\hat{\phi} L}{\beta_1} \right) - \frac{\lambda L^2}{\beta_2} x \left( \frac{(1 - \hat{\phi}) L}{\beta_2} \right) + \alpha L q (e^{-q\hat{\phi}} + e^{-q(1-\hat{\phi})}) \\ &= -\lambda L^2 (c - 1 - \alpha) \left( \frac{1}{\beta_1} + \frac{1}{\beta_2} \right) + \alpha L \left( \left( q - \frac{\lambda L}{\beta_1} \right) e^{-q\hat{\phi}} + \left( q - \frac{\lambda L}{\beta_2} \right) e^{-q(1-\hat{\phi})} \right). \end{aligned}$$

Since total labor participation  $L$  is bounded below by [Lemma 4](#), the upper bound is concave for all  $\lambda > \Lambda_1$  big enough.

The first derivative

$$\begin{aligned} \left. \frac{\partial \tilde{\pi}}{\partial \phi} \right|_{\phi_0 = \hat{\phi}} &= \left. \frac{\partial \bar{\pi}}{\partial \phi} \right|_{\phi_0 = \hat{\phi}} = -\underline{N}_1(\hat{\phi}, \hat{\phi}, L) - \alpha L \underline{M}_1(\hat{\phi}, \hat{\phi}) \\ &= L \left[ x \left( \frac{\hat{\phi} L}{\beta_1} \right) - x \left( \frac{(1 - \hat{\phi}) L}{\beta_2} \right) - \alpha \left( e^{-q\hat{\phi}}(1 - q\hat{\phi}) - e^{-q(1-\hat{\phi})}(1 - q(1 - \hat{\phi})) \right) \right] \\ &\stackrel{(3)}{=} \alpha L \left[ e^{-q\hat{\phi}} - e^{-q(1-\hat{\phi})} - e^{-q\hat{\phi}}(1 - q\hat{\phi}) + e^{-q(1-\hat{\phi})}(1 - q(1 - \hat{\phi})) \right] \\ &= \alpha L \left[ q\hat{\phi}e^{-q\hat{\phi}} - q(1 - \hat{\phi})e^{-q(1-\hat{\phi})} \right] \end{aligned}$$

has the same sign as  $\mathcal{Q}(\hat{\phi}) - q$ . This allows us to conclude: If  $q > \mathcal{Q}(b)$ , there exists  $\delta > 0$  small enough such that  $q > \mathcal{Q}(b + \delta)$ . By [Theorem 1](#), there exists  $\Lambda_2$  big enough such that the unique stable steady state has composition  $\hat{\phi} \in [b, b + \delta)$ , and hence surplus is locally decreasing at the steady state for all  $\lambda > \max\{\Lambda_1, \Lambda_2\}$ . If  $q < \mathcal{Q}(b + \delta)$ , the opposite is true. Since this holds for all levels  $L > 0$ , it also holds for the upper envelope, implying  $\phi^* \leq \hat{\phi}$ .

Finally, total labor participation is optimal at any steady state, since the supply constraints in [Equation \(3\)](#) ensure that the first order condition [\(8\)](#) holds at any steady state  $(\hat{\phi}, \hat{L})$ .  $\square$

**Proof of [Theorem 6](#).** To prove (a), we establish existence of an optimum by citing the Extreme Value Theorem and noting that  $\tilde{\pi}$  is continuous over the bounded set  $[0, 1] \times [0, \beta_1 + \beta_2]$ . We proceed by ruling out all optimal labor force compositions other than  $(0.5, b)$ .

To start, we rule out any  $\phi < 0.5$  through symmetry. In such a situation, the simplified expression for long-run surplus [\(16\)](#) implies

$$\tilde{\pi}(1 - \phi, L) - \tilde{\pi}(\phi, L) = \int_{\phi L}^{(1-\phi)L} \left[ F^{-1} \left( 1 - \frac{s}{\beta_1} \right) ds - F^{-1} \left( 1 - \frac{s}{\beta_2} \right) \right],$$

which is strictly positive whenever  $\phi < 0.5$ . In other words, at any total labor size  $L$ , composition  $1 - \phi$  dominates  $\phi$ .

To further restrict the range of optimal labor force compositions, we proceed similar to the proof of [Theorem 1](#). We define an auxiliary function  $Z : [0, 1] \rightarrow \mathbb{R}$  as

$$\begin{aligned} Z(\phi) &= b - \phi - (1 - \phi)bF(c - 1 - \alpha\mu(\phi) + \alpha(1 - \phi)M(\phi)) \\ &\quad + \phi(1 - b)F(c - 1 - \alpha\mu(1 - \phi) - \alpha\phi M(\phi)) \end{aligned}$$

for  $M(\phi) = (1 - \phi)\mu'(1 - \phi) - \phi\mu'(\phi)$ . [Lemma 5](#) shows that the roots of  $Z$  contain the optimal interior composition  $\phi$ . [Lemmata 6](#) and [7](#) establish sufficient conditions such that  $Z(\phi) < 0$  for all  $\phi \in [b, 1)$ , and  $Z(0.5) = (b - 0.5)(1 - F(c - 1 - \alpha\mu(0.5))) >$

0 for  $\alpha > 0$  small enough.

The last case to rule out is a homogeneous labor force with  $\phi = 1$ . To do so, we look at the sign of the partial derivative (see [Lemma 5](#) for derivation details)

$$\lim_{\phi \rightarrow 1} \tilde{\pi}_\phi(\phi, L) = L \left[ F^{-1} \left( 1 - \frac{L}{\beta_1} \right) - \lim_{p \rightarrow 1^-} F^{-1}(p) + \alpha(\mu(1) - \mu(0) + \mu'(1)) \right]. \quad (15)$$

If talent is unbounded above,  $\lim_{p \rightarrow 1^-} F^{-1}(p) = \infty$ , no (bounded) mentoring gains ever justify excluding (unbounded) talent. Conversely, if there exists a maximal talent  $\hat{x}$ , the optimal homogeneous labor size satisfies the first order condition  $\tilde{\pi}_L(1, L^*) = 0$  and hence  $F^{-1} \left( 1 - \frac{L^*}{\beta_1} \right) = c - 1 - \alpha\mu(1)$ . For any  $\alpha \in \left( 0, \frac{c-1-\hat{x}}{\mu'(1)} \right)$ , long-term surplus is then locally decreasing at the boundary,  $\tilde{\pi}_\phi(1, L^*) < 0$ , which rules out an optimum at  $\phi = 1$ .

To prove (b), we show that no steady state admits a labor force composition  $\phi \in [0.5, b]$ . Indeed, since  $F(c-1) < 1$ , a minimal mass of either group invests into education regardless of mentoring. As before, we can identify mixed steady state compositions through the roots of  $S$ . And as before,  $S(\phi) > 0$  for any  $\phi \in [0.5, b]$ . Indeed, neither of the relevant parts of the proofs for [Lemma 1](#) and [Theorem 1\(d\)](#) rely on any functional form assumptions, and translate directly to this setting after replacing  $x^{-1}$  by  $1 - F$ . Without persistent intervention, long term surplus is thus bounded above by

$$\max \{ \tilde{\pi}(\phi, L) \mid \phi \in [0, 0.5] \cup [b, 1], L \in [0, \beta_1 + \beta_2] \},$$

which is strictly dominated by the intervention identified above. □

**Lemma 5.** *Let  $Z : [0, 1] \rightarrow \mathbb{R}$  be defined as in [Theorem 6](#). Then  $Z(\phi^*) = 0$  at any interior optimal labor force composition  $\phi^*$ .*

*Proof.* For constant interventions of composition  $\phi$  and total size  $L$ , and after a change of variables  $x \mapsto s = \beta_i(1 - F(x))$ , long-run surplus simplifies from

Equations (6) and (9) to

$$\begin{aligned} \tilde{\pi}(\phi, L) = & \int_0^{\phi L} F^{-1}\left(1 - \frac{s}{\beta_1}\right) ds + \int_0^{(1-\phi)L} F^{-1}\left(1 - \frac{s}{\beta_2}\right) ds \\ & + L(1-c) + \alpha L(\phi\mu(\phi) + (1-\phi)\mu(1-\phi)). \end{aligned} \quad (16)$$

At any interior optimum, two first order conditions hold jointly:

$$\begin{aligned} 0 = \tilde{\pi}_\phi = & L \left[ F^{-1}\left(1 - \frac{\phi L}{\beta_1}\right) - F^{-1}\left(1 - \frac{(1-\phi)L}{\beta_2}\right) \right. \\ & \left. + \alpha(\mu(\phi) - \mu(1-\phi) - M(\phi)) \right], \end{aligned} \quad (17)$$

$$\begin{aligned} 0 = \tilde{\pi}_L = & \phi F^{-1}\left(1 - \frac{\phi L}{\beta_1}\right) + (1-\phi)F^{-1}\left(1 - \frac{(1-\phi)L}{\beta_2}\right) \\ & + 1 - c + \alpha(\phi\mu(\phi) + (1-\phi)\mu(1-\phi)). \end{aligned} \quad (18)$$

Multiplying (17) by  $\frac{1-\phi}{L}$  and adding it to (18) is equivalent to

$$\frac{\phi L}{\beta_1} = 1 - F(c - 1 - \alpha\mu(\phi) + (1-\phi)\alpha M(\phi)). \quad (19)$$

Similarly, multiplying (17) by  $\frac{\phi}{L}$  and subtracting it from (18) is equivalent to

$$\frac{(1-\phi)L}{\beta_2} = 1 - F(c - 1 - \alpha\mu(1-\phi) - \phi\alpha M(\phi)). \quad (20)$$

The function  $Z$  is obtained by subtracting  $\phi(1-b)$  times (20) from  $(1-\phi)b$  times (19). At the optimal composition, it must therefore equal  $(1-\phi)b\frac{\phi L}{\beta_1} - \phi(1-b)\frac{(1-\phi)L}{\beta_2} = 0$ .  $\square$

**Lemma 6.** *Assume that there exists  $k > 0$  and  $\underline{x} < \bar{x}$  such that  $F(\underline{x}) > 0$ ,  $F(\bar{x}) < 1$ ,  $F' > 0$  over  $[\underline{x}, \bar{x}]$  and*

$$\frac{1 - F(x_1)}{1 - F(x_2)} > \frac{k}{k + x_1 - x_2} \quad \forall \underline{x} < x_2 < x_1 < \bar{x} \quad (x^{\text{NP}})$$

$$\alpha(\mu(\phi) - \mu(1 - \phi) - M(\phi)) < k \left( \frac{\phi(1 - b)}{b(1 - \phi)} - 1 \right) \quad \forall \phi \in [b, 1) \quad (\mu^{\text{NP}})$$

$$c - 1 - \alpha\mu(\phi) + \alpha(1 - \phi)M(\phi) \in [\underline{x}, \bar{x}] \quad \forall \phi \in [0, 1] \quad (A1^{\text{NP}})$$

Then  $Z(\phi) < 0$  for all  $\phi \in [b, 1)$ .

*Proof.* Assume  $\phi \in [b, 1)$  and note that  $Z(\phi) < 0$  is equivalent to

$$\frac{1 - F(x_1)}{1 - F(x_2)} > \frac{(1 - \phi)b}{(1 - b)\phi}, \quad (21)$$

for  $x_1 = c - 1 - \alpha\mu(1 - \phi) - \alpha\phi M(\phi)$  and  $x_2 = c - 1 - \alpha\mu(\phi) + \alpha(1 - \phi)M(\phi)$ , since (A1<sup>NP</sup>) ensures that the denominator is positive. The difference  $x_1 - x_2$  is bounded above by Property ( $\mu^{\text{NP}}$ ). We now proceed by cases:

- If  $\phi = b$ , this upper bound equals zero, and hence  $x_1 < x_2$ . The strict monotonicity of  $F$  implies that  $\frac{1 - F(x_1)}{1 - F(x_2)} > 1$ , and hence (21).
- If  $x_1 \leq x_2$ , monotonicity of  $F$  similarly implies that  $\frac{1 - F(x_1)}{1 - F(x_2)} \geq 1$ , and  $1 > \frac{(1 - \phi)b}{(1 - b)\phi}$  for all  $\phi > b$ .
- If  $x_1 > x_2$ , we apply the assumptions to conclude that

$$\frac{1 - F(x_1)}{1 - F(x_2)} \stackrel{(x^{\text{NP}})}{>} \frac{k}{k + \alpha(\mu(\phi) - \mu(1 - \phi) - M(\phi))} \stackrel{(\mu^{\text{NP}})}{>} \frac{(1 - \phi)b}{(1 - b)\phi}. \quad \square$$

**Lemma 7.** *If there exists  $\underline{x} < c - 1 < \bar{x}$  and  $v > 0$  such that  $F(\bar{x}) < 1$  and  $F' < v$  over  $[\underline{x}, \bar{x}]$ , then there exists  $k > 0$  such that Property ( $x^{\text{NP}}$ ) holds. Furthermore, if both  $\mu'$  and  $\mu''$  are bounded over  $[0, 1]$  and Property (10) is satisfied, then there exists  $\alpha > 0$  small enough such that Properties ( $\mu^{\text{NP}}$ ) and (A1<sup>NP</sup>) also hold.*

*Proof.* The upper bound on  $F'$  extends to the secant slope,  $F(x_1) - F(x_2) = \int_{x_2}^{x_1} F'(x)dx < v(x_1 - x_2)$ . Letting  $k = \frac{1 - F(\bar{x})}{v} > 0$ , it thus follows that  $1 - F(x_1) > vk$  and

$$\frac{1 - F(x_1)}{1 - F(x_2)} = \frac{1}{1 + \frac{F(x_1) - F(x_2)}{1 - F(x_1)}} > \frac{1}{1 + \frac{v(x_1 - x_2)}{1 - F(x_1)}} > \frac{1}{1 + \frac{1}{k}(x_1 - x_2)} = \frac{k}{k + x_1 - x_2}.$$

For the second claim, let  $\mathcal{M}(\phi) = (1 - \phi)(\mu(\phi) - \mu(1 - \phi) - M(\phi))$ . Condition (10) ensures that  $\mathcal{M}(b) < 0$ . The bounds on  $\mu'$  and  $\mu''$  translate to an upper bound on  $\mathcal{M}'$ , i.e.  $\mathcal{M}'(\phi) \leq m_1$  for all  $\phi \in (b, 1)$ . It then follows that for any  $\alpha \in \left(0, \frac{k}{\max\{m_1, 1\}b}\right]$  and any  $\phi \in [b, 1)$ ,

$$\mathcal{M}(\phi) = \mathcal{M}(b) + \int_b^\phi \mathcal{M}'(y) dy \stackrel{(10)}{<} m_1(\phi - b) \leq \frac{k}{\alpha b}(\phi - b) = \frac{(1 - \phi)}{\alpha} k \left( \frac{\phi(1 - b)}{b(1 - \phi)} - 1 \right).$$

Multiplying both sides by  $\frac{\alpha}{1 - \phi}$  yields **Property ( $\mu^{\text{NP}}$ )**. Similarly, the bounds on  $\mu$  and  $\mu'$  imply bounds on  $\mu(\phi) + (1 - \phi)M(\phi) \in (-m_2, m_2)$  over  $[0, 1]$ . **Property (A1<sup>NP</sup>)** holds for any  $\alpha \in \left(0, \min\left\{\frac{c-1-\underline{x}}{m_2}, \frac{\bar{x}-c+1}{m_2}\right\}\right)$ , and that interval is nonempty since  $c - 1 \in (\underline{x}, \bar{x})$ .  $\square$

## 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.
- Beaman, Lori, Raghendra 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.
- 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.
- 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.
- 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.
- Commission for Employment Equity**, “Annual Report 2016-2017,” 2017.

- 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.
- 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.
- 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.
- 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.
- 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.