MENTORING AND THE DYNAMICS OF AFFIRMATIVE ACTION

By

Michèle Müller-Itten and Aniko Öry

November 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2112

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

MENTORING AND THE DYNAMICS OF AFFIRMATIVE ACTION*

Michèle Müller-Itten† and Aniko Öry‡

November 14, 2017

Abstract

We study the evolution of labor force composition when mentoring is more effective within members of the same socio-demographic type. Typically, multiple steady states exist. Some completely exclude juniors of one type. Even a mixed steady state tends to over-represent the type that is dominant in the population. In contrast, the efficient labor force balances talent recruitment against mentoring frictions. It may even under-represent the dominant type and typically calls for persistent government intervention. This contrasts with the public discourse around temporary affirmative action. We consider specific policy instruments and show that hiring quotas can induce equilibrium employment insecurity.

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, Kirsten Cornelson, Thomas Gresik, Ben Hermalin, Sandile Hlatshwayo, Lisa Kahn for insightful discussions and comments. Financial support from NSF (grant # SES 1227707) is gratefully acknowledged.

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

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

1 Introduction

Empirical studies have shown that mentoring relationships are stronger between members of the same demographic group.[1] 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.[2] The achievement differences arise early on and manifest themselves through different education choices.[3] Given the importance of mentoring for human capital development, and the implications on labor market outcomes, the diversity of the pool of potential mentors can have persistent social welfare implications.

This paper provides a theoretical framework to study inter-generational mentoring, and we use it to understand the need and the optimal design of affirmative action policies. The most surprising insight is that persistent intervention in favor of the minority can be optimal. This policy does not vanish in magnitude over time, nor phase out like policies aimed at 'leveling the playing field'. It is qualitatively different than those motivated by restoration and fairness, arguments at the heart of the initial affirmative action movement.[4] Our model calls for such persistent intervention when minority talent recruitment is critical for surplus maximization. Subsidies then allow agents to internalize the social benefits that arise from

[1] 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.

[2] The empirical literature documents three important channels: First, teachers of similar demographics boost student performance and graduation rates (Bettinger and Long, 2005; Dee, 2007). Notably, the performance gap between white and underrepresented minority students drops by 20-50 percent when taught by a minority instructor (Fairlie et al., 2014). 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). Second, faculty bias generates hurdles for minority students. Teachers disproportionately fail to identify talented minority students (Card and Giuliano, 2016). A teacher of different race is 33% more likely to perceive a student as inattentive when compared to an own-race teacher evaluating the same student (Dee, 2005). Third, the composition of political leadership affects labor market outcomes, since officials invest more into infrastructure that is directly relevant to the needs of their own genders (Chattopadhyay and Duflo, 2004).

[3] Bayer and Rouse (2016) show that 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.

[4] In what is often seen as the founding moment of affirmative action, President Johnson famously argued that "You do not take a person who, for years, has been hobbled by chains and liberate him, bring him up to the starting line of a race and then say, 'you are free to compete with all the others,' and still justly believe that you have been completely fair." (Johnson, Lyndon B. "To Fulfill These Rights." Commencement Address at Howard University, Washington DC, June 4, 1965.)

efficient minority mentoring. These benefits do not disappear in the long run, and so neither do the subsidies.

We consider a continuous-time overlapping generations model in which every person belongs to one of two types (1 and 2), 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.

The cost of investment is a function of innate ability (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.[5] The key parameters in our model are talent concentration, mentor capacity and majority dominance. Talent concentration is largest in high skilled labor markets such as the market for doctors, lawyers or professors. Zero talent concentration means that all individuals have equal talent. We assume no ex-ante differences in talent distribution across the two groups. Mentor capacity is given by how many mentees one mentor can care for at a point in time. This parameter depends on the type of mentor interaction: We expect it to be high for classroom instruction but low for one-on-one coaching situations. Finally, majority dominance refers to the share of majority types in the overall population: It is roughly equal to 0.5 in the case of gender and larger in the case of race in the United States where around 76.9% of the population is white.[6]

The trade-off at the core of our analysis is between an optimal use of innate talent and reduced mentoring frictions. A mixed workforce may fail to match juniors with a same-

[5] We assume a fixed mentoring technology (which we elaborate on in the conclusion), 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. Along these lines, the academic senate at Berkeley writes: "Information about grants, the politics of a department, or opportunities for participation in professional activities are often shared with younger colleagues upon meeting them in social settings to which some groups may have less access. Lacking power and access to the occasions in which power is shared, some groups are offered different opportunities for participation in the academic world" (Committee on the Status of Women and Ethnic Minorities (SWEM), "Guide to Faculty Advancement and Promotion at U.C. Berkeley," *Academic Senate, Berkeley Division*, 2012, page 18).

[6] See U.S. Census Bureau QuickFacts, as retrieved from https://www.census.gov/quickfacts/fact/table/US/RHI125216 on 09/29/2017.

type mentor (particularly if mentor capacity is low or minority mentors are in short supply) but attracts the most talented juniors of both types. A homogeneous workforce minimizes mentoring frictions (all juniors and mentors are of the majority type), but high entry barriers for the minority keep even its most talented juniors out of the workforce.

We show that there exist multiple steady states; these may involve either a completely homogeneous or a mixed workforce, depending on the values of the three parameters mentioned above. The mixed steady state generates higher social surplus if mentor capacity or talent concentration is sufficiently high. Due to competition and free entry, firms do not internalize that today's hires affect tomorrow's candidate pool. Thus, the market outcome is in general inefficient and temporary affirmative action policies can increase total surplus if they are sufficiently strong to assure convergence to the desirable steady state.[7]

The *optimal* policy intervention need not be temporary however. Indeed, we show under what conditions a planner intervenes persistently in favor of a minority group. This is in contrast to much of the public discourse that sees affirmative action as a temporary policy tool to correct a historical imbalance.[8] An efficient strategy achieves this goal quickly, the reasoning goes, and thereby renders itself obsolete in a relatively short amount of time. We show that this is only generally true if the fraction of the two groups in the population is balanced, which is true in the case of gender, but not in the case of race or other minorities.

Perhaps surprisingly, the optimal persistent policy may even enforce a fraction of minority mentors in the labor force which is *larger* than the fraction of minority juniors in the population. The reason lies in the large positive mentoring externality that minority workers generate for future students. Indeed, when the minority is *over*-represented in the labor force, it actually means that the marginal minority student has *less* innate talent than the

[7] If the intervention is too small, the economy will eventually fall back to the undesirable steady state. 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 above this threshold. This 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.

[8] Consider for example the debate following the Supreme Court decision to uphold a ban on affirmative action in Michigan's public university admissions (Kahlenberg et al., 2014). Most arguments center around the question of whether or not historical racial discrimination has been overcome.

marginal majority student. Despite this, efficiency dictates that such a composition should be implemented indefinitely.

We also compare different concrete policy instruments, such as educational subsidies like scholarships and workplace hiring quotas under varying assumptions on wage determination and identify the winners and losers of different policies. In our framework, the optimal workforce can be achieved using educational subsidies that are budget neutral in the long run. Hiring quotas are equally effective as long as the competitive environment allows for type-dependent 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 hiring quotas among educated majority workers who are excluded from the labor market. To minimize this job insecurity, our theory suggests that under a hiring quota, efficient wages for minorities should be *higher* than for majority workers.[9]

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.[10] Under statistical discrimination, employment quotas may actually reinforce negative stereotypes against certain groups (Coate and Loury, 1993).[11] Quotas are completely ineffective in altering beliefs when agents infer their personal success probability from their own type's employment history as in Chung (2000).[12] We complement this discussion by allowing for

[9]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).

[10]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.

[11]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.

[12]The finding in Milkman et al. (2015) that professors are less likely to respond to students who are not

tangible mentoring complementarities, and show that this sheds a much more positive light on affirmative action policies. This is in line with Ben-Porath (1967) who views human capital as being produced using innate talent and other inputs (which could be mentoring). Becker and Tomes (1979); Restuccia and Urrutia (2004); Herskovic and Ramos (January 24, 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. The goal of this paper is to provide a new perspective that can help policy makers to trade off costs and benefits of affirmative action and to evaluate different policies.

The structure of our model is similar to that of Athey et al. (2000), who study the intertemporary promotion decision in long-lived firms. As in their model, we assume that the marginal benefit of an additional mentor from the same group is decreasing, but by parametrizing the model in terms of relevant industry variables, we are able to provide comparative statics and offer policy recommendations.

2 Model

2.1 Educational Investment

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 type $i = 1$ (majority) and type $i = 2$ (minority). Upon birth, each junior has a one-time opportunity to invest into costly education, and thereafter seeks employment in a competitive and unsaturated labor market. For the sake of simplicity, we assume that education is instantaneous and the only determinant of later productivity.[13]

white and male, could potentially also be explained by this theory. However, their study is about prospective students anonymously reaching out to mentors, while our theory is probably more applicable in a context with personal contact.

[13] 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.

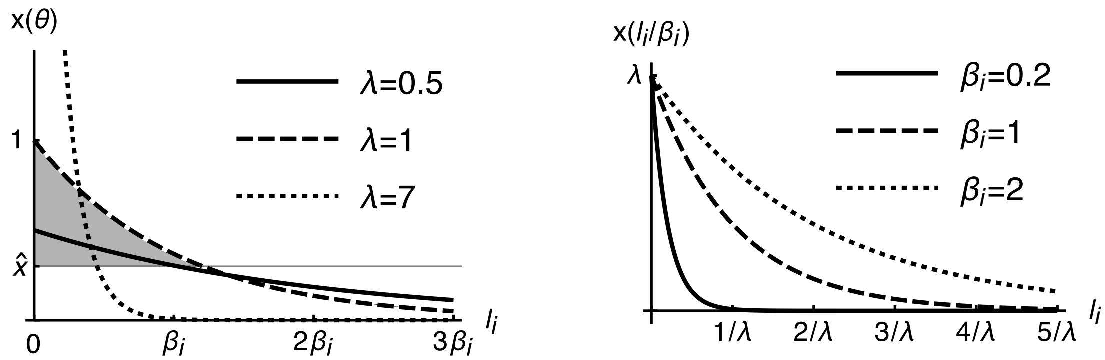
A junior invests into education if his expected lifetime earnings outweigh the cost of education. Earnings are determined through market forces: We assume that each unit mass of educated senior workers contributes one unit to a firm's profit flow, uneducated workers contribute nothing. Life expectancy follows a standard exponential distribution with parameter 1. Assuming free entry of firms, this ensures that the expected lifetime earnings of each educated senior equal $w_i = 1$ in an unregulated labor market. The cost of education, c, is reduced both by the junior's individual talent and the type-specific strength of mentoring. The next two paragraphs specify how we model innate talent and mentoring boosts.

Talent is identically distributed across types. We consider the decreasing *talent function* $x : [0, \infty) \to [0, \lambda]$ given by

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

Extending Athey et al. (2000), we allow the two junior pools to be of unequal *scale* $\beta_1 \geq \beta_2 > 0$ by assuming that a mass $\beta_i\theta$ of group i has talent larger than $x(\theta)$, for all ranks $\theta > 0$. Put differently, there are β_1 type-1 juniors for every β_2 type-2 juniors at every talent level x. All qualitative results are driven by the ratio $b = \beta_1/\beta_2 \geq 1$, which we refer to as *majority dominance*. In either pool, a fraction $\frac{\hat{x}}{\lambda}\left(1 - \ln\left(\frac{\hat{x}}{\lambda}\right)\right)$ of total talent is concentrated among juniors with talent $x \geq \hat{x}$ (shaded region in Figure 1a). The parameter λ therefore measures talent *concentration* – it is particularly large for specialized education that requires rare skills, such as doctors, lawyers or professors.

Educated seniors form mentoring relationships with incoming juniors. There are four main features that we want to capture: First, mentorship is a scarce resource, limited by the ratio of seniors to juniors. Second, mentoring is more effective if the senior and junior are of the same type. Third, there are matching frictions due to limited mentor availability within a certain sub-discipline, geographical area or individual school. Fourth, there are decreasing returns to scale from mentorship for an individual junior. To capture these features parsimoniously, we consider the limit (as $n \to \infty$) of a discrete matching market



(a) talent concentration λ

(b) pool scale β_i

Figure 1: Talent distribution as a function of parameters λ and β_i.

with nL_i seniors of group i and nl juniors (of any type). Each senior-junior link exists with probability $\frac{q}{nl}$,[14] where *mentor capacity* q denotes the average number of juniors per mentor. A junior of type i enjoys a mentorship boost of 1 if and only if he or she is being mentored by at least one type-i mentor. By the Law of Rare Events, the number of same-type mentors per junior can be approximated by the Poisson distribution as n grows. The probability of finding a same-type mentor, and hence the expected *mentorship boost* for a junior of type i, is therefore equal to

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

where $\Lambda_i = L_i/l$ is the ratio of group-i seniors to juniors. Figure 2 illustrates the effect of mentor capacity on the mentorship boost function. This parameter is determined by industry specific skills and norms: Classroom instruction allows one mentor to reach more juniors than individual coaching. Decreasing trends in the time invested in mentoring (DeLong et al., 2008) will also affect market dynamics through this channel.

For simplicity, we assume that there are no complementarities between talent and mentorship boost. As a result, a type-i junior of talent x_i invests in education if and only if

$$c - x_i - \alpha\mu\left(\frac{L_i}{l}\right) \leq w_i, \tag{1}$$

[14]For n large enough, the link probabilities are non-degenerate, $0 < \frac{q}{nl} < 1$.

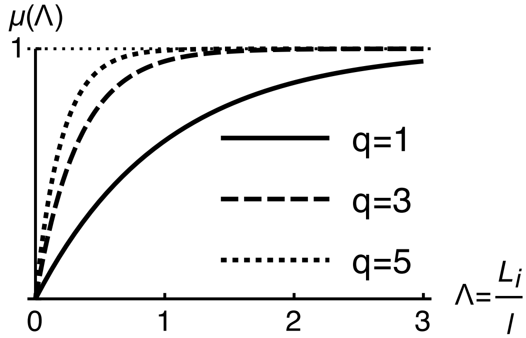


Figure 2: Mentorship boost is increasing in mentor availability Λ and mentor capacity q.

where $\alpha > 0$ measures the relative importance of mentoring versus innate skill. The individual rationality constraint (1) reveals interesting dynamics: First, mentors are increasingly busy during periods of labor growth ($L_i \ll l$), resulting in a smaller boost for juniors. Second, adding seniors of either group (weakly) lowers the educational costs for *all* juniors, ceteris paribus. Indirectly however, seniors of the opposite type 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.

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:

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

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

2.2 Labor Force Evolution

Combining individual rationality constraints (1) across juniors, the equilibrium conditions for junior investment $\boldsymbol{l} = (l_1, l_2)$ given mentoring pool $\boldsymbol{L} = (L_1, L_2)$ are

$$\begin{cases} c - x\left(\frac{l_1}{\beta_1}\right) - \alpha\mu\left(\frac{L_1}{l_1+l_2}\right) = w_1 \quad \text{or} \quad l_1 = 0 \text{ and } c - \lambda - \alpha\mu\left(\frac{L_1}{l_1+l_2}\right) > w_1, \\ c - x\left(\frac{l_2}{\beta_2}\right) - \alpha\mu\left(\frac{L_2}{l_1+l_2}\right) = w_2 \quad \text{or} \quad l_2 = 0 \text{ and } c - \lambda - \alpha\mu\left(\frac{L_2}{l_1+l_2}\right) > w_2, \end{cases} \tag{2}$$

where wage $w_i = 1$ for the case without market intervention.[15] This system of equations admits a unique solution $\boldsymbol{l}$ that responds intuitively to changes in wage.

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

Proof. See Appendix A. □

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 $\boldsymbol{L}(t) = (L_1(t), L_2(t)) \in \mathbb{R}^2_+$ therefore evolves according to the differential equation

$$\dot{\boldsymbol{L}}(t) = \boldsymbol{l}(t) - \boldsymbol{L}(t), \tag{3}$$

with $l_i(\tau) \geq 0$ denoting the mass of type-i juniors born at time τ who invest in education. 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.[16]

Definition. *The economy is in a* steady state *when $\dot{\boldsymbol{L}} = 0$ or, equivalently, $\boldsymbol{l} = \boldsymbol{L}$. A steady state $\hat{\boldsymbol{L}} \in \mathbb{R}^2_+$ is* stable *whenever a small perturbation does not affect the long-term*

[15] In Section 4.3, we discuss how quotas or scholarships affect wages differently for each type.

[16] This same concept is sometimes referred to as "asymptotically stable" in the literature.

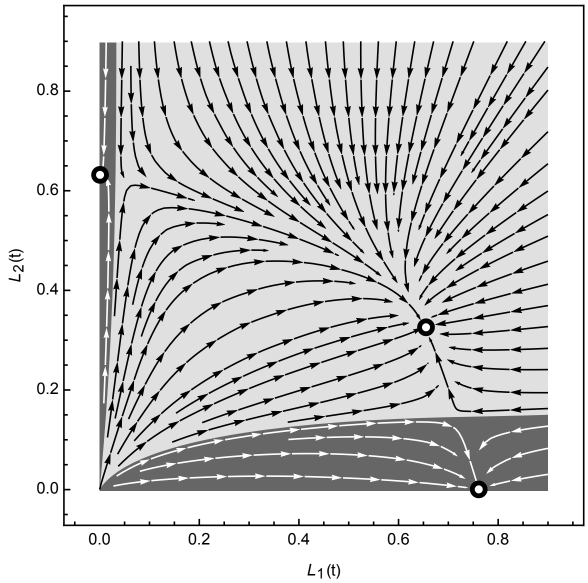


Figure 3: Labor force evolution $\dot{\boldsymbol{L}}$ at any $\boldsymbol{L}$. The three stable steady states are highlighted along with their respective convergence regions.

convergence, i.e., when there exists $\varepsilon > 0$ *such that for all* $\mathbf{L}(0) \in \mathbb{R}^2_+$ *with* $\|\mathbf{L}(0) - \hat{\mathbf{L}}\| < \varepsilon$, $\lim_{t\to\infty} \mathbf{L}(t) = \hat{\mathbf{L}}$.

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 $\boldsymbol{L}$. This is possible because junior investment (2) depends on calendar time only through the senior labor force $\boldsymbol{L}(t)$. Exponential life expectancy ensures the same for labor force evolution (3). Finally, the Hartman-Grobman Linearization theorem offers an equivalent notion of stability that also omits time dependence.

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

Figure 3 visually represent the labor force evolution $\dot{\boldsymbol{L}}$ as a vector map.[17] The Linearization Thoerem formalizes the idea that a steady state is stable if and only if all surrounding arrows point towards it. In this case, there are three stable steady states: Two are *ho-*

[17]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 $q = 3$.

mogeneous $(L_1, 0), (0, L_2) \in \mathbb{R}^2_+$ and one is *mixed*. The starting pool of mentors $\boldsymbol{L} \in \mathbb{R}^2_+$ determines towards which of these the labor force will ultimately converge.

3 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* $L = L_1 + L_2$.[18] 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 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 mentoring 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

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

(c) Under Assumption (A1$^+$) and for sufficiently high mentor capacity q or talent concentration λ, there exists a unique stable mixed steady state with ϕ arbitrarily close to majority

[18] Indeed, the former is a sufficient for the steady state analysis since Equation (2) uniquely determines the corresponding L (see Lemma 3 in the appendix).

dominance 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 type-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 type may exceed this mentoring boost. 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 (1). If either mentor capacity or talent concentration is large, $\hat{x}_i$ hardly responds to differences in mentor availability since $\mu(\Lambda) \to 1$ as $q \to \infty$ and $x'(x^{-1}(\hat{x})) \to -\infty$ as $\lambda \to \infty$. As a result, the mixed stable steady state is unique and approaches majority dominance b (claim c). More generally, we show that *all* stable mixed steady states involve an over-representation of the majority under (mSS$^+$). Property (mSS$^+$) can be interpreted as a lower bound on c, q or α.[19]

To prove Theorem 1 formally, 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 $(0, b]$.

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

Lemma 2. *Let $S : [0,1] \to \mathbb{R}$ be given by*

$$S(\phi) = (b - \phi)\ln(\lambda) - (1-\phi)b\ln(c - 1 - \alpha\mu(\phi)) + \phi(1-b)\ln(c - 1 - \alpha\mu(1-\phi)). \tag{4}$$

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

Proof. The one-to-one correspondence between steady-state composition and the roots of S stems directly from Equation (2) at a steady state $\boldsymbol{L} = \boldsymbol{l} = (\phi L, (1-\phi)L)$,

$$(1-\phi)bx^{-1}(c - 1 - a\mu(\phi)) = \frac{\phi(1-\phi)}{\beta_1 + \beta_2}L = \phi(1-b)x^{-1}(c - 1 - a\mu(1-\phi)) \Leftrightarrow \frac{S(\phi)}{\lambda} = 0.$$

Property (mSS) ensures that at least one of the terms in the first equation is positive, and hence total labor supply $L > 0$. Stability owes to the Hartman-Grobman Linearization Theorem and is relegated to Appendix A. □

In addition to these existence results, Lemma 2 also allows us to anticipate the impact of parameter changes on the labor composition in a mixed stable steady state. Figure 4 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, $b \downarrow$,

(b) talent becomes more concentrated, $\lambda \uparrow$, or

(c) mentor capacity increases, $q \uparrow$,

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

Proof. See Appendix A. □

It is not surprising that the composition of the population (captured by majority dominance 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 λ 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 type 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 type-imbalance.[20] Theorems 1 and 2 suggest that this can be due to two reasons. Either, we might be moving from a homogenous steady state 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 type-imbalance is larger at the top in the steady state, this would imply that the differences in mentor technology are more important (from a positive standpoint) than those in skill concentration. In the next section, we study the related normative question and identify what the composition *should be* in order to maximize total surplus.

4 Welfare

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

[20] 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.

investments,

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

where $\boldsymbol{L} \in \mathbb{R}^2_+$ denotes the senior labor force and $\boldsymbol{l} \in \mathbb{R}^2_+$ the new recruits. The shaded area in Figure 1a visually represents the surplus generated by a type with marginal talent $\hat{x} = x(\theta/\beta_i)$. 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.

4.1 Welfare maximizing 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 ($\boldsymbol{L} = \boldsymbol{l}$), total surplus can be written as

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

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 redirect the economy through a temporary intervention.

Theorem 3 (Optimal Steady State)**.** *For sufficiently large mentor capacity q or high talent concentration λ, 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 short-term course correction in favor of the underrepresented type.

Combining the insight from Theorems 1 and 3 yields another important takeaway: Temporary intervention does not achieve a workforce that accurately reflects the diversity in the population. Indeed, under Property (mSS$^+$), the minority (with $\beta_2 < \beta_1$) remains *under*represented 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 *over*represents the minority for this precise reason.

4.2 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* $\dot{\boldsymbol{L}}^*(\boldsymbol{L})$.[21] 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 ϕ and total participation L. In Section 4.3, 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

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

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 *over*represents 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* ϕ^* *depends on mentor capacity* q *and talent concentration* λ*:*

(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* $\lim_{q\to\infty} \phi^* = b$*.*

(d) For large enough λ*, the optimal labor force is more balanced than in the unique mixed stable steady state,* $\phi^* < \hat{\phi}$ *if and only if* $q > \mathcal{Q}(b) = \frac{1}{2b-1} \ln\left(\frac{b}{1-b}\right)$*.*

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

Proof. See Appendix A. □

To illustrate the individual claims, Figure 4 plots the optimal and steady state labor force composition against mentor capacity. A stable mixed steady state exists for $q \geq 2.74$; it

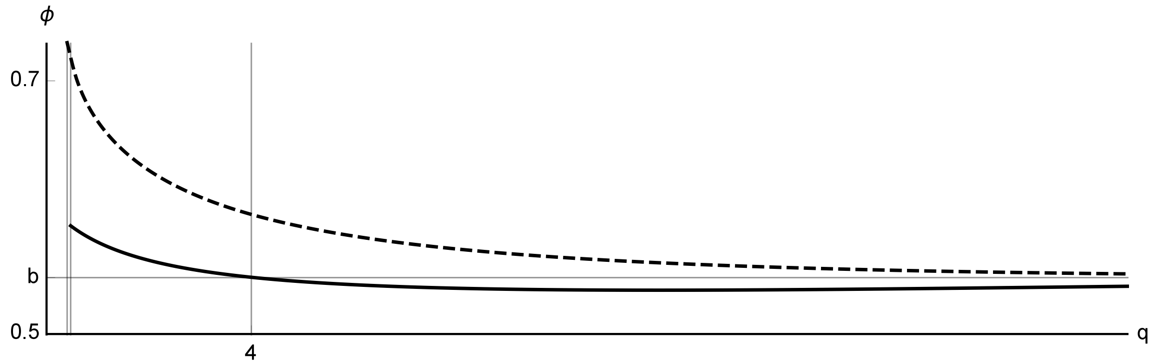


Figure 4: Optimal (solid) and steady state (dashed) labor force composition as a function of mentor capacity.

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 dominance 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.

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.[22] In particular, there is no reason to assume that affirmative action policies render themselves obsolete by virtue of their own success, contrary to supreme-court majority ruling which argued that "race-conscious ad-

[22]Indeed, the result shows one particular intervention that – while not necessarily fully optimal – generates positive social surplus, and hence dominates a 'laisser-faire' regime.

missions policies must be limited in time" and expects them to disappear within 25 years.[23]

Finally, Theorem 4 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.

4.3 Policy Instruments

In the previous sections, we have shown conditions under which the policy maker wishes to ensure junior investment $\boldsymbol{l}^*$ rather than the myopic $\boldsymbol{l}$ given senior labor force $\boldsymbol{L}$ 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: Type-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 type-specific fellowships and tuition hikes.[24] Let $\boldsymbol{\Delta} \in \mathbb{R}^2$ denote such a transfer schedule where Δ_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 $\boldsymbol{\Delta}$ satisfies

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

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

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

[24] 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 (2). 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.

order condition

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

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

Labor Force Quotas. Alternatively, the policy maker can restrict the recruitment decisions of firms by setting caps on the type 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 types among all educated new hires.[25] We call a quota ϕ^* *binding* at $\boldsymbol{L}$ if it forces the firm to recruit more minority types than they would myopically. Formally, if $\boldsymbol{l}$ denotes the solution to Equation (2) under wages $w = 1$, ϕ^* 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** type-i individuals by $\bar{\boldsymbol{l}} \geq \boldsymbol{l}^*$ respectively.

There are limits to the compositions that can be implemented through a quota. Notably, the zero-profit condition for firms equates the marginal cost of an educated hire to the added

[25] 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.

productivity,

$$\phi^* w_1 + (1 - \phi^*) w_2 = 1, \tag{9}$$

Since no workers accept negative wages, this limits the set of feasible market wages.[26] Quotas generate desired investment levels only if educational investment is sufficiently responsive to wage differences. To avoid lengthy digressions, we study situations where education retains a positive net cost even to the most able individual under perfect mentoring,

$$c - \lambda - \alpha > 0. \tag{A2}$$

Note that if $\alpha \leq 1$ or $\lambda < 1$, this follows from Property (hSS) or (A1$^+$) 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.[27]

When wages are determined solely through market forces, all educated workers find employment.[28] Workers' individual rationality (Lemma 1) and the zero profit-condition (9) uniquely determine the instantaneous cohort $\bar{\boldsymbol{l}} = \boldsymbol{l}^* = (\hat{\phi}, 1 - \hat{\phi})l \in \mathbb{R}^2_{>0}$ and equilibrium earnings $w_1 < 1 < w_2$. In particular, a binding quota *raises* minority and *depresses* majority earnings relative to the unconstrained market. 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 agree on the optimal total participation.[29]

In some industries however, social or legal pressure prohibits paying unequal wage to employees in the same position.[30] Without differential hiring bonuses, the zero-profit condi-

[26] To avoid anticipatory investment, we assume that quotas are unanticipated, and that constrained firms complement wages by (positive or negative) *hiring bonuses* $\boldsymbol{w} - \boldsymbol{1}$ upon initial appointment.

[27] Without minority mentors, a quota shuts down the labor market if mentoring is critical ($\alpha \gg 0$). Minority investment then remains zero even the largest possible wages; consequently, firms can hire nobody. In contrast, for any $L_2 > 0$, Assumption (A1) ensures that any wage $w_2 \geq 1$ attracts minority students as long as majority investment l_1 is small enough. Low wages $w_1 \to 0$ have the desired effect by Assumption (A2).

[28] Any oversupply of educated type-i workers would drive their lifetime earnings to zero. Anticipating this, none of them would invest into education under Assumption (A2).

[29] Restating hiring bonuses as subsidies $\boldsymbol{\Delta} = \boldsymbol{1} - \boldsymbol{w}$, the zero-profit condition (9) is equal to budget balance.

[30] This is the stated rationale behind the 2014 presidential memorandum titled 'Advancing Pay Equality Through Compensation Data Collection' (Presidential Memorandum, 79 Fed.Reg. 20751 (Nov.04, 2014),

tion (9) forces the market wage to 1. However, a binding quota caps the demand for type-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 (2) 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.[31] Majority workers waste their own resources on an ex-post worthless education and dilute mentoring efficiency for everybody else. This 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.

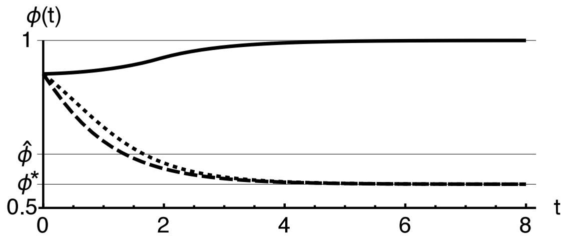
Theorem 5 (Policy Instruments)**.** *The policy maker can implement the optimal long-run labor force* $\boldsymbol{L^*}$ *through educational subsidies that are budget neutral in the long run. Under Assumption (A2), a hiring quota* $\phi^* = \frac{L_1^*}{L_1^*+L_2^*}$ *implements the same* $\boldsymbol{L}^*$ *if and only if type-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. □

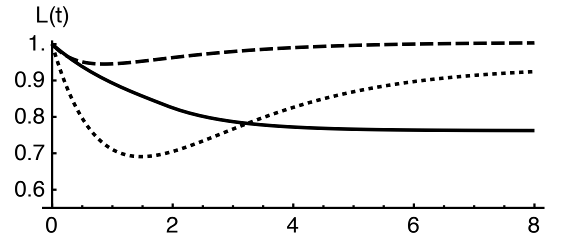
To illustrate, Figure 5 plots the evolution of key labor market variables under a hiring quota. The optimal long-run composition ϕ^* is more diverse than the mixed steady state $\hat{\phi} > \phi^*$. The starting value $\boldsymbol{L}(0)$ is such that an unconstrained economy (solid line) converges to a homogeneous steady state (panel a). Imposing quota ϕ^* 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 π_C well below

www.federalregister.gov/d/2014-08448). Firms also have internal incentives to avoid type-dependent 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).

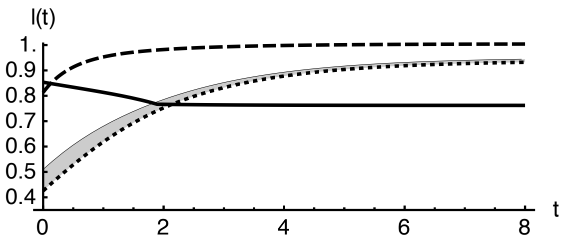
[31] The student body is more balanced than in the unregulated case if and only if $w_1 < 1$ (Lemma 1). Equilibrium job insecurity therefore accompanies any binding quota.



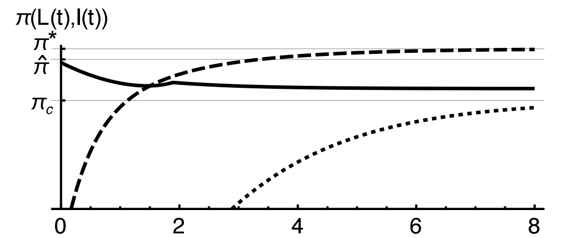
(a) Labor force composition $\phi(t)$



(b) Total labor force participation $L(t)$



(c) Total recruits $l(t)$



(d) Instantaneous surplus $\pi(\boldsymbol{L}(t), \boldsymbol{l}(t))$

Figure 5: Labor force evolution starting at $\boldsymbol{L} = (0.8, 0.2)$ 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.

the optimal level π^*, and even below the laisser-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 Conclusion

Opponents of positive discrimination often criticize the *persistence* of affirmative action policies. The very fact that proponents still advocate for lower admission standards for minorities, the argument goes, demonstrates their ineffectiveness in bringing about lasting change.[32] Our analysis highlights the shortcomings of such an inference in the face of mentoring complementarities. In that environment, minority workers possess a rare and valuable skill: They can significantly lower educational costs for future minority students. Majority workers do not have the same skill, and mentorship for future majority students

[32]By way of example, consider Sacks, David and Peter Thiel, "The Case against Affirmative Action." *Stanford Magazine*, September/October 1996.

is already in abundant supply. The surplus-maximizing scholarships, or the market wages, remunerate minorities for this valuable skill.

We show that the social externality may be so large that it warrants a long-term *over*-representation of the minority. Such a policy implies that minorities are not merely compensated for a current lack of suitable mentors, but actively courted through incentives that are unavailable to majority students of equal talent. This goes against another rhetoric frequently used to decry affirmative action: Fairness.[33] Students who are equal in every aspect save for race, they argue, should be treated equally. There may be room to compensate minorities for lacking support structures, but any policy that goes beyond 'leveling the playing field' amounts to discrimination against the majority. Under mentoring complementarities however, the equality of the two students is fictional: The minority student possesses rare mentoring skills that do more for future talent recruitment than those of his majority twin.

Important 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-type mentoring skills, tend to decrease access hurdles in a way similar to minority scholarships. Policy makers have been fostering such mentoring for minority youth,[34] and the most successful diversity programs are exactly those that increase cross-type exposure.[35] These approaches chip away at the same obstacles

[33] Leonhardt, David. "Rethinking Affirmative Action." *New York Times*, October 13, 2012.

[34] 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`).

[35] 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.

that we study, albeit from an angle that is not considered here. Third, our welfare measure is focused on total surplus, but the suggested policy interventions generate both winners and losers. Much like information rents, high-talent minority students receive additional resources despite the fact that education has been viable for them all along. Majority students are forced to bear the cost of these scholarships through increased tuition or lower wages; and under sticky wages, quotas may even cause employment insecurity for educated majority workers.

We do not want this paper to be read in isolation. Affirmative action has many important consequences and we study just one aspect. However, we hope to add an awareness of *mentoring complementarities* to this discussion. On the most basic level, it is this: 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 type-specific mentoring ability according to the shortness of its supply and its impact on future surplus.

A Additional Proofs

Proof of Lemma 1. Consider a total student body of size l, and let

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

The equilibrium conditions (2) 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 type i; type-j enrollment $\ell_j(l)$ drops due to the increase in l. $\square$

Lemma 3. *At a homogeneous steady state with type 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 (2). 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 4. *The function $g(\phi) = \ln(c - 1 - \alpha\mu(\phi))$ is decreasing and convex under Assumption (A1$^+$). As $q \to \infty$, $g(\phi) \to \ln(c - 1 - \alpha)$ and $g'(\phi) \to 0$ pointwise for all $\phi \in (0,1)$. Finally, $\partial g(\phi)/\partial q = \frac{\phi}{q} g'(\phi)$ is increasing in ϕ.*

Proof. Negative monotonicity follows immediately from monotonicity of μ 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$^+$).

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

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 ϕ by convexity of g. $\square$

Supplement to the proof of Lemma 2. 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 ϕ is stable whenever $S'(\phi) < 0$.

The equation of motion $\dot{\boldsymbol{L}} = \boldsymbol{l}(\boldsymbol{L}) - \boldsymbol{L}$ 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 $\frac{\partial \dot{\boldsymbol{L}}}{\partial \boldsymbol{L}}$ have a negative real part. For simplicity, we first rewrite the equations in (2) as

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

where $g(\phi)$ is as in Lemma 4.

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

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

After simplification using Lemma 3, the characteristic polynomial $\left|\frac{\partial \dot{\boldsymbol{L}}}{\partial \boldsymbol{L}} - \gamma I\right|$ is proportional to $F(\gamma) = A\gamma^2 + B\gamma + C$ with

$$\begin{aligned} A &= \frac{\lambda L}{\beta_2}\left(-(1-b)\frac{\lambda L}{\beta_2} + (1-b)(1-\phi)g'(1-\phi) + b\phi g'(\phi)\right) < 0 \\ B &= -\frac{2}{1-b}\left[S'(\phi) + \frac{S(\phi)}{b-\phi}\right]^2 - \left[S'(\phi) + \frac{S(\phi)}{b-\phi}\right]\left[(1+2\phi)g'(1-\phi) + (3-2\phi)\frac{b}{1-b}g'(\phi)\right] \\ &\quad - \left[(1-b)\phi g'(1-\phi)^2 + \frac{b^2}{1-b}(1-\phi)g'(\phi)^2\right] \\ C &= \frac{\lambda L}{\beta_2}\left[S'(\phi) + \frac{S(\phi)}{b-\phi}\right]. \end{aligned}$$

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 $B < 0$. □

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 type has a steady-state mentorship boost of $\mu(1)$, which determines labor supply through Equation (2). 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 $\boldsymbol{l} = \boldsymbol{L} = \big(\hat{\phi}\hat{L}, (1-\hat{\phi})\hat{L}\big) \in \mathbb{R}^2_+$, where the individual cost-benefit analyses in Equation (2) simplify to

$$c - x\left(\frac{\hat{\phi}\hat{L}}{\beta_1}\right) - \alpha\mu(\hat{\phi}) = 1 \qquad \text{and} \qquad c - x\left(\frac{(1-\hat{\phi})\hat{L}}{\beta_2}\right) - \alpha\mu(1-\hat{\phi}) = 1. \tag{10}$$

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.[36]

The limiting behavior of g implies the existence of a unique stable mixed steady state near b for q or λ large enough. We first show existence, and start with the case $q \to \infty$ first. It follows from Lemma 4 that

$$S(b) = (1-b)b(g(1-b) - g(b)) > 0 \tag{11}$$

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

$$S'(b) = (1-b)g(1-b) + bg(b) - \ln\lambda - (1-b)b(g'(1-b) + g'(b)) \xrightarrow{q\to\infty} k = \ln(c-1-\alpha) - \ln(\lambda) \overset{(A1)}{<} 0.$$

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

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

[36] 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 4). Together with $S(b) = (1-b)b(g(1-b) - g(b)) \geq 0$, this also implies a zero over $[b, 1)$.

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

$$S(b+\delta) \le S(b) + \delta S'(b) - \frac{k}{2}\delta < \frac{k}{2}\delta - \frac{k}{2}\delta = 0 \qquad \forall q \ge Q_1. \tag{12}$$

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 λ and unbounded below. Hence, there exists $\Lambda_1 > 0$ such that

$$S(b+\delta) < 0 \qquad \forall \lambda > \Lambda_1. \tag{12'}$$

By continuity, Equation (11) and either (12) or (12') imply that $S(\phi)$ crosses 0 downwards at some $\phi \in (b, b+\delta)$. By Lemma 2, 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$^+$). For large q or λ, we now partition $[0,1]$ into five intervals and pin down the location stable steady state in one of them.

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

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

This is unbounded below in λ and converges to $-\ln(\lambda) + \ln(c-1-\alpha) < 0$ as $q \to \infty$. Consequently, there exist Λ_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,

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

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 ε, $S(\phi)$ is positive over $I_2 = [\phi_0, \frac{1-b}{2}]$,

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

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

Over-representation of the majority can be shown in two steps based on Lemma 2. First, S admits no root over $\phi \in (0.5, b)$ since

$$
S(\phi) = (b-\phi)(\ln(\lambda) - g(\phi)) + \phi(1-b)(g(1-\phi) - g(\phi)),
$$

where $g(\phi) = \ln(c-1-\alpha\mu(\phi)) \overset{(A1)}{<} \ln\lambda$ is decreasing. For $1-\phi < \phi \leq b$, both terms are nonnegative and the second is strictly positive. Second, for $0 < \phi < 0.5 \leq b$, it is possible to write

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

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$^+$) ensures concavity of $h(\phi) = g(1-\phi) - g(\phi)$ over $(0, 0.5)$ and implies the last inequality. Together, these inequalities imply that any steady state ($S(\phi) = 0$) with type-1 in the minority ($b - \phi > 0.5 - \phi > 0$) must have positive slope $S'(\phi) > 0$, and hence be unstable. Symmetry rules out the last candidate steady state $\phi = 0.5$ for $\beta_1 > \beta_2$. □

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 4) 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 $S(b) = (1-b)b\big(g(1-b) - g(b)\big) > 0$ by Lemma 4. Continuity of S then implies the existence of a downward-crossing (and by Lemma 2 a stable steady state) over $(b, \phi) \subseteq (0.5, \phi)$. □

Proof of Theorem 3. Define the single-type 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 β_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.[37] As $\lambda \to \infty$, most terms disappear and $\pi_i(m) \to \beta_i$ pointwise for all values of m. Similarly, as $q \to \infty$, $\mu(\phi) \to 1$ and hence $\pi_i(\mu(\phi)) \to \pi_i(1)$ pointwise for all values of ϕ.

In Theorem 1, we establish the existence of a stable steady state arbitrarily close to composition b for large enough q or λ. The surplus of that steady state eventually exceeds that of a homogeneous workforce since

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

and

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

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. □

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

Lemma 5. *Long-term surplus $\tilde{\pi}(\phi, L)$ attains its maximum (ϕ^*, L^*) at some $\phi \in [0.5, 1]$ and*

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

$L^* \in (0, \bar{L}]$ *for* $\bar{L} = \frac{\beta_1+\beta_2}{c-1-\alpha\mu(1)}$. *Under Assumption (A1$^+$) 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 $\rho(L_i, \beta_i, L) = \beta_i\big(1 - e^{-\lambda\frac{L_i}{\beta_i}}\big) + L_i\,(1 - c + \alpha\mu(L_i/L))$ is strictly supermodular in its first two arguments since $\frac{\partial^2 \pi_i}{\partial L_i \partial \beta_i} = \lambda^2/\beta_i^2 L_i e^{-\lambda\frac{L_i}{\beta_i}} > 0$. 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 dominant type, $\phi^* \geq 0.5$.

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

$$\begin{aligned}
\tilde{\pi}(\phi, L) &< \beta_1 + \beta_2 + \phi L \underbrace{\big(1 - c + \alpha\mu(\phi)\big)}_{<1-c+\alpha\mu(1)<0} + (1-\phi)L \underbrace{\big(1 - c + \alpha\mu(1-\phi)\big)}_{<1-c+\alpha\mu(1)<0} \\
&< \beta_1 + \beta_2 + \tilde{L}\big(1 - c + \alpha\mu(1)\big) < 0.
\end{aligned}$$

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 type, 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 $q > Q$. □

Proof of Theorem 4. The existence of a surplus-maximizing labor force with positive supply and type-1 majority follows directly from Lemma 5. 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 $q \to 0$ by Property (hSS) and to $c - 1 - \lambda - \alpha < 0$ as $q \to \infty$ whenever (A1) holds for any q. They are ordered as $q_0 > \underline{q}$ since μ is strictly increasing in ϕ 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 mentoring 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 ρ is as in Lemma 5. 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 $\phi = \phi_0$,

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

Long-term surplus can be written as

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

Replacing M and N with $\underline{M}$ and $\underline{N}$ yields a quadratic upper bound $\overline{\pi}$, for any ϕ_0.

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

$$\left.\frac{\partial^2 \overline{\pi}}{\partial \phi^2}\right|_{\phi_0 = b} = \underbrace{-(\lambda L)^2 \left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right) e^{-\frac{\lambda L}{\beta_1 + \beta_2}}}_{\text{negative and independent of } q} + \alpha L \underbrace{q(e^{-qb} + e^{-q(1-b)})}_{\to 0 \text{ as } q \to \infty}, \tag{13}$$

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

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

is also negative for $q > Q_2$ large enough.[38]

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

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

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 5 and Assumption (A1$^+$), 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 $\frac{\partial \underline{N}}{\partial \phi}(b, L) \equiv 0$ for any $L > 0$. Moreover, by

[38]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.

Equation (13) 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

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

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 $\lim_{q\to\infty}(c_1 + c_2 q)e^{-c_3 q}$ 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 $L^* = \arg\max_L\{\tilde{\pi}(\phi, L)\}$. It follows that

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

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

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

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

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

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

(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 (2) implies for the leading coefficient

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

Since total labor participation L is bounded below by Lemma 5, the upper bound is concave for all $\lambda > \Lambda_1$ big enough.

The first derivative

$$\begin{aligned}
\left.\frac{\partial \tilde{\pi}}{\partial \phi}\right|_{\phi_0=\hat{\phi}} = \left.\frac{\partial \overline{\pi}}{\partial \phi}\right|_{\phi_0=\hat{\phi}} &= -\underline{N}_1(\hat{\phi}, \hat{\phi}, L) - \alpha L \underline{M}_1(\hat{\phi}, \hat{\phi}) \\
&= L\left[x\left(\frac{\hat{\phi}L}{\beta_1}\right) - x\left(\frac{(1-\hat{\phi})L}{\beta_2}\right) - \alpha\left(e^{-q\hat{\phi}}(1-q\hat{\phi}) - e^{-q(1-\hat{\phi})}(1-q(1-\hat{\phi}))\right)\right] \\
&\overset{(2)}{=} \alpha L\left[e^{-q\hat{\phi}} - e^{-q(1-\hat{\phi})} - e^{-q\hat{\phi}}(1-q\hat{\phi}) + e^{-q(1-\hat{\phi})}(1-q(1-\hat{\phi}))\right] \\
&= \alpha L\left[q\hat{\phi}e^{-q\hat{\phi}} - q(1-\hat{\phi})e^{-q(1-\hat{\phi})}\right]
\end{aligned}$$

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 Λ_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 $\phi^* \lesseqgtr \hat{\phi}$.

Finally, total labor participation is optimal at any steady state, since the supply constraints in Equation (2) ensure that the first order condition (8) holds at any steady state $(\hat{\phi}, \hat{L})$. □

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

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

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

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

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

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

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

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.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *Econometrica*, 2004, *72* (5), 1409–1443.

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

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

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

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

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

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

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

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

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

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

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

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," *Unpublished working paper*, January 24, 2017. 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.

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.

Milkman, Katherine L, Modupe Akinola, and Dolly Chugh, "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations.," *Journal of Applied Psychology*, 2015, *100* (6), 1678.

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.