Underrepresentation, Quotas and Stigma: A dynamic argument for reform*

Malin Arve and Justin Valasek†

September 4, 2023

Abstract

The tension between increased representation and stigma is central to the debate on whether to use quotas to address underrepresentation in high-profile professions. We address this trade-off using a dynamic model of career selection where juniors value both the identity and prestige of their mentors (seniors). A preference for homophily causes persistence of underrepresentation, even if discrimination is eliminated, suggesting intervention is needed. However, if an abrupt quota causes a high level of stigma, then underrepresented juniors of high talent will select out of the profession, causing a persistence of stigma. Encouragingly, we show that gradual reform—while introducing some stigma in the short term—enables a transition to a stigma-free steady state with equal representation in the long term. We discuss the implications of our analysis for commonly-used measures to increase representation.

Keywords: Affirmative Action, Quotas, Mentorship, Identity, Gender, Adverse Selection.
JEL Classification Codes: D62, E24, I2, J15, J16, J24.

*We would like to thank Sarah Auster, Rahul Deb, Thomas Giebe, Hans Grönqvist, Bard Harstad, Steffen Huck, Botond Kőszegi, Rachel Kranton, Konuray Mutluer and seminar participants at BI: Bergen, the 2022 EUI Alumni Conference, and the Nordic Theory Workshop for their comments and suggestions.

†Norwegian School of Economics (NHH). Contact e-mail: malin.arve@nhh.no and justin.valasek@nhh.no

1 Introduction

The persistent underrepresentation of women and minority groups in many high-skilled professions is a well-documented phenomenon, and has led to an active policy debate over how to tackle underrepresentation.[1] While certain pundits and scholars have argued against the use of quotas and affirmative action on the basis that it introduces a stigma of lower quality among underrepresented professionals, it is unclear whether this argument outweighs the clear benefit of increasing representation and diversity in high-skilled professions. Importantly, the stigma argument only considers a static perspective and largely ignores the dynamic arguments for increased representation. A large empirical literature has documented the importance of role models in the decision of labor-profession entrants to pursue a specific career, suggesting that affirmative action today may result in a more representative workforce tomorrow (see Porter and Serra, 2020, and Riise et al., 2022 for an overview). Therefore, any potential stigma associated with quotas may be transitory as a profession moves towards equal representation.

To address the interplay between increased representation and stigma, we analyze identity-based hiring quotas in a formal dynamic model and are thus able to provide additional structure to this debate. We show that stigma is not an argument against quotas to increase representation. Stigma, however, can have important implications for how quotas are implemented. Specifically, we show that quotas

[1] See for example Auriol et al. (2022) for global evidence from the academic profession for economists and Wallon et al. (2015); European Commission (2019) for a overview of the policy debate.

are necessary to transition to a steady state with equal representation. However, if a quota results in a large shift in the relative perceived quality (stigma), then it can cause high-talent juniors of the underrepresented type select out of the profession, resulting stigma and real quality differences between identity types. Encouragingly, we also show that there always exists a gradual path of reform that enables a transition to a stigma-free steady state with equal representation in the long term.

We follow the example of Athey et al. (2000) and Müller-Itten and Öry (2022) and consider a overlapping generations (OLG) model where mentorship plays a key role in the development of quality, and where potential career entrants (juniors) value both the identity and quality of their mentors.[2] To account for discrimination, we also introduce the concept of prestige as a separate object from quality: quality is private information and is productive in the sense that juniors are more likely to realize high quality if their mentors (seniors) have high quality, while prestige is a publicly observable signal that is correlated with quality. Discrimination occurs when, on average, seniors of different identity types with the same quality have different levels of prestige.

Our analysis of this model shows that given a preference for homophily—i.e. juniors prefer mentors of their own identity type—underrepresentation is persistent. This is the result of a cycle where, due to the relative lack of seniors of the same identity type, juniors of the under-represented identity disproportionately select out of the profession. This cycle causes persistent underrepresentation at

[2] Mentorship is an important factor in many different career areas such as politics, law and academia. Moreover, identity-homophily is a well-documented fact in some fields of academia (Hilmer and Hilmer, 2007; Gaule and Piacentini, 2018) and in the judiciary branch (Battaglini et al., 2022).

the senior level and illustrates the need for policy intervention.

Interestingly, we also find discrimination is endogenously "hidden" in our model. While both groups of seniors have the same average level of prestige, the average quality is therefore higher among seniors of the identity group that faces discrimination. This means that juniors of the identity type facing discrimination realize a higher average quality due to the higher quality of their mentors. However, at a steady state, this higher average quality is precisely offset by the direct impact of discrimination. Therefore, the dynamics of the observable metric of prestige are the same with and without discrimination: despite discrimination, an equal proportion of juniors are hired as seniors from each identity type and each group has the same average prestige.

We subsequently consider policy solutions to address underrepresentation. While it is unclear whether or not eliminating discrimination is a realistic policy in our setting, we nonetheless show that it is ineffective at eliminating underrepresentation. That is, while eliminating discrimination will cause underrepresented juniors to enter the profession at a higher rate, it is not enough to ensure that the profession will transition to equal representation due to the persistence of underrepresentation.

Next we consider a 1:1 quota on juniors as a way of correcting for underrepresentation at the senior level. However, we find that a quota on juniors does not solve the problem of underrepresentation due to an adverse selection problem: while a quota on juniors mechanically equalizes representation at the junior level, it does not increase the number of high-talent juniors of the underrepresented type, and therefore may not lead to an increase in the number of *seniors* of the underrepresented type.

Therefore, in our main analysis, we consider quotas for hiring seniors of the under-represented type as a direct approach to addressing underrepresentation. Quotas introduce stigma by lowering the average prestige of seniors in the under-represented type. However, it is important to note that quotas do not necessarily cause lower average quality in the underrepresented type due to discrimination. Indeed despite that fact that empirical studies have shown that quotas do not result in lower quality in the target group,[3]affirmative action and quotas can negatively impact the *perceptions* of quality (e.g. Heilman et al., 1992; Coate and Loury, 1993; Fang and Moro, 2011; Leslie, 2014).

We find that even if quotas do not result in lower quality in the seniors of the underrepresented type, stigma may impact quality through entry decisions of career entrants. Specifically, while identity-based hiring quotas at the senior level mechanically address underrepresentation, the level of stigma and quality of the implemented steady state depends on the dynamic structure of the quotas. That is, quotas can either result in a transition to a steady state with equal prestige of seniors of both types, or result in a transition to a steady-state where high-talent juniors of the underrepresented type select out of the profession, resulting in both persistent stigma and a lower average quality of seniors of the historically underrepresented type.

[3]For example Besley et al. (2017) show that a quota actually *increased* the average quality of politicians in Sweden, a country considered to be one of the most egalitarian in the world. This, suggests that hiring quotas may in practice serve to address discrimination.

The transition to an unequal steady-state can occur since, with a preference for homophily, juniors disproportionately value the quality of seniors of their identity type. Therefore, if a quota causes a large enough decrease in the prestige of seniors of the underrepresented type, then high-talent juniors of the underrepresented type will disproportionately select out of the profession, causing a transition to a steady state with unequal quality. Our analysis therefore points to a gradually increasing quota on seniors of the underrepresented type as a way to ensure a transition to a stigma-free steady state with equal representation.

In addition to the literature on role-models discussed above, our research contributes to the theoretical literature on affirmative action and underrepresentation (see Fershtman and Pavan, 2021 for an overview). Our work is most closely related to Athey et al. (2000) and Müller-Itten and Öry (2022), who study quotas in the context of juniors who value the identity-composition of the mentor pool and find that quotas may be required to maintain equal or optimal representation. Arguments for quotas to address underrepresentation are also presented in Siniscalchi and Veronesi (2020) and Carvalho and Pradelski (2022) using alternative models of underrepresentation based on a mechanism of, respectively, self-image bias and in-group norms. We expand on this research by explicitly modeling discrimination and accounting for the fact that juniors' career decisions may also depend on the perceived quality-composition of the mentor pool, which allows us to address the important trade-off between representation and stigma that is often central to the debate surrounding affirmative action and quotas. This innovation leads to our novel insight that the dynamics of quotas matter: in contrast to previous research, we highlight that the speed of reform is crucial because it can impact whether the profession converges to equal quality, or to a steady state where high-talent juniors of the underrepresented type select out of the profession.

Lastly, we discuss the implications of our analysis for measures that have been proposed or implemented for addressing underrepresentation. Our results suggest that the "cascade model"—a quota at each level of seniority that is equal to the level of representation at the level below—can be counterproductive since it may result in a lower perceived quality of underrepresented seniors relative to *both* a more gradual transition and to an immediate transition to equal representation. In contrast, a preference for underrepresented seniors in cases of equal quality avoids the problem of the cascade model, but can lead to a cycle where representation is increased in one period and reduced in the next. Therefore, this model may require an occasional "nudge" to keep the profession on the path towards equal representation.

2 Theoretical Framework

We consider an OLG setup where each agent lives for two periods, and there is an overlapping population of juniors and seniors in each period. In the first period each agent is a career entrant and can apply for a junior position in a given profession which includes on-the-job mentoring by a senior colleague. For shortness of the exposition, we will refer to the two levels of positions as juniors and seniors. Conditional upon being hired after the junior period, he or she becomes a senior. In each period, the profession consists of a continuous population of seniors of mass 1. Each senior has the capacity to mentor $\lambda > 1$ juniors, implying that a maximum of λ juniors can be trained in each period. There is a continuous population of career entrants (potential juniors) of each identity and talent type of size N, where N is arbitrarily large.[4] Identity and talent is described in more detail below. For simplicity, we first introduce the notation without a time subscript.

Types: Over the life-cycle, each agent i is characterized by a five-dimensional type $(I_i, q_i, \pi_i, Q_i, o_i)$ where we refer to I_i as the agent's identity-type, q_i, as the talent of the agent as a junior (i.e., in the first period), π_i is the prestige of the agent as a senior, Q_i the quality of the agent as a senior (i.e., in the second period) and o_i the value of the agent's outside option.

The identity-type space is binary and each agent i has an identity $I_i \in \{A, B\}$ that is observable and constant over time. This identity-type can for example be the agent's gender or ethnicity. We denote by $\mathbf{M}^I$ (respectively $\mathbf{m}^I$) the set of senior (resp. juniors) of identity-type I.

Junior talent is binary, and we denote a junior's talent by $q_i \in \{h, l\}$. Juniors' talent is private information, and relates to their ability to realize high quality/prestige as seniors (we detail the production function of prestige and quality below). In the first stage, juniors have an outside option which is valued at $o_i = o_{q_i}$ with $o_h > o_l$.

Seniors' prestige and quality are continuous variables that are imperfectly cor-

[4] This assumption assures that our results are not driven by a limited supply of juniors of a given identity type.

related at the individual level. Senior prestige is public information, while senior quality is unobserved. Conceptually, prestige can be thought of as individual reputation, while quality determines the ability to produce juniors with high quality/prestige. Importantly, by modeling prestige and quality separately, we are able to introduce discrimination as a biased belief regarding the relationship between prestige and quality—we remain agnostic as to whether the profession's objective is to maximize prestige, quality, or some combination of the two.

Timing and Choices: At the beginning of each period, agents first decide whether to enter the profession and apply for a position which includes on-the-job mentoring (e.g. a PhD program or a junior program at a law firm). All juniors who apply have an equal probability of being hired. There is a fixed application cost c for applying. This may represent actual monetary costs or the cost of specialized training (e.g. GRE prep, law school, etc.). We use the notation $\hat{a}_i = 1$ to denote that i applies for a job, and $\hat{a}_i = 0$ if i does not apply. Since the number of junior positions is fixed at λ, not all applicants will become juniors. We use the notation $a_i = 1$ to denote that i obtained a junior position, and $a_i = 0$ that i did not.

Then, conditional upon entry, juniors receive mentoring and realize prestige and quality $\{\pi_i, Q_i\}$. We do not explicitly model the mentoring process (see an earlier version of the paper, **?**, that models the matching of juniors to seniors); however, we introduce a production function of quality/prestige that accounts for the key features of the matching process of juniors to mentors when juniors both value quality and have a preference for homophily. Lastly, a mass 1 of juniors are hired as seniors for $t+1$. Since only π_i is observable at the individual level and the profession has a strict preference for quality/prestige, the hiring rule consists of an endogenous prestige cutoff, π_L, above which all seniors are hired (we formally introduce identity quotas in Section 4).

Payoffs: Agents who do not enter training have utility equal to the value of their outside option. Juniors' utility is represented by the following function:

$$u^I_{q_i}(\pi_i, \Pi^I, \nu^I), \tag{1}$$

where $u^I_{q_i}$ is increasing in π_i and Π^I, and $\nu^I = m^I - \lambda M^I$ is a measure of the relative number of juniors and seniors of type I. That is, we model a preference for homophily as follows: a junior of type I receives a negative utility if the mentoring capacity of seniors of identity type I is lower than the number of juniors of type I. Specifically:

$$\partial u^I_{q_i}(\pi_i, \Pi^I, \nu^I)/\partial \nu^I \begin{cases} = 0 & \text{if } \nu^I \leq 0, \\ < 0 & \text{if } \nu^I > 0. \end{cases}$$

While we do not model the matching of juniors to mentors in this version of our paper, conceptually this negative utility is associated with the probability of being matched with a mentor of a different type, which is assumed to be positive if the set of juniors of one type is larger than the mentoring capacity of seniors of that type.

Production of prestige and quality: At the individual level, prestige and quality are two related, but distinct features, and we model π_i as a noisy measure of Q_i. Prestige is a public reputation and is the only metric used in the job market for seniors. Quality, which is not observed publicly, is productive in mentoring; i.e. it determines the ability to produce juniors with high quality/prestige.

Specifically, absent discrimination, all juniors draw prestige π_i from a left-censored uniform distribution with a probability mass of $(1 - f(q_i, \bar{Q}^I, m_h))$ at 0, and a probability mass of $f(q_i, \bar{Q}^I, m_h)$ distributed uniformly between $[0,1]$. That is, the probability that a junior draws a strictly positive prestige (between $(0,1]$) is equal to:

$$f(q_i, \bar{Q}^I, m_h), \tag{2}$$

where the function $f(\cdot)$ maps $(q_i, \bar{Q}^I, m_h) \to [0,1]$. Next, to capture π_i as a noisy measure of Q_i we assume that $Q_i = 0$ when $\pi_i = 0$, but that $Q_i = \pi_i + \varepsilon_i$ when $\pi_i > 0$, where ε_i is drawn from a uniform distribution with support $[-\varepsilon, \varepsilon]$.

Additionally, we allow for discrimination of seniors with identity A, which we model as a leftward shift in the distribution for prestige. That is, juniors of identity type A draw prestige from a distribution with mass at $-d$, and $f(q_i, \bar{Q}^A, m_h)$ distributed uniformly between $[-d, 1-d]$. Quality, however, is not affected by discrimination in the sense juniors of identity A have $Q_i = \pi_i + d + \varepsilon_i$. That is, given $\bar{Q}^A = \bar{Q}^B$, juniors of each identity type have the same average quality, but juniors of type A on average realize lower prestige. We assume that juniors are naive about discrimination in the sense that they hold the belief that $d = 0$.

The function $f(l, \bar{Q}^I, m_h)$ has the following features: First, to simplify our analysis, we assume that high talent juniors drive quality and that for low talent juniors $f(l, \bar{Q}^I, m_h) = 0$. [5] Next, we assume that quality is productive, and that $\partial f(h, \bar{Q}^I, m_h)/\partial \bar{Q}^I > 0$. Lastly, we assume that $f(h, \bar{Q}^I, m_h)$ is decreasing in the number of other high talent juniors in the profession ($\partial f(h, \bar{Q}^I, m_h)/\partial m_h < 0$). That is, there is a competition effect of high talent.

The rationale for the competition effect is based on the matching to mentors. While we do not model the matching process explicitly in this version of the paper (again, see **?**), we find it reasonable to assume that the highest prestige seniors are more likely to match up with high-talent juniors, and that these matches have the highest probability of generating prestige/quality. Therefore, from the perspective of a high-talent junior, more high talent increases the competition for mentorship with the highest-prestige seniors, which we capture with the assumption that $\partial f(h, \bar{Q}^I, m_h)/\partial m_h < 0$.

Lastly, note that $f(q_i, \bar{Q}^I, m_h)$ is a function of average quality of seniors of the junior's own identity, $\bar{Q}^I$. This captures homophily and the fact that juniors are more likely to receive mentoring from a senior with the same identity type.[6] In a previous version of the analysis, we microfound this assumption by explicitly modeling a matching process of juniors to seniors in a setting where juniors have a preference for homophily.

Equilibrium and steady state:

[5] Our main results also hold under the assumption that low-talent juniors realize positive senior quality and where $f(l, \Pi^I, m_h) < f(h, \Pi^I, m_h)$ for all Π^I. However, the assumption that all low-talent juniors realize $Q_i = 0$ allows us to characterize the dynamics of $\bar{Q}$ in a relatively straightforward matter, and to more cleanly illustrate our main results.

[6] It is also sufficient to assume that the average quality of seniors of their own type has a strictly larger impact on the probability that juniors realize positive quality/prestige.

We consider symmetric period equilibria as a set of probabilities of applying, $\sigma_q^I = \Pr(\hat{a}_i = 1|q,I)$, that maximize the expected utility given $\{\mathbf{M}^I\}$. Note that while we explicitly model the decision to enter the industry and apply for a position, our analysis will focus on characterizing the size of the sets of juniors that enter in equilibrium: $\{m_h^A, m_h^B, m_l^A, m_l^B\}$, where $m_{q_i}^I = |\mathbf{m}_{q_i}^I|$. Therefore, much of the machinery and notation we introduce in this section will operate in the background (i.e. the appendix).

In our dynamic analysis, three important metrics that function as state variables are the size of the different sets of senior identity types, which we denote with $M^I = |\mathbf{M}^I|$, average senior quality for identity type I, $\bar{Q}^I$, and average senior prestige, Π^I. We define a steady state as $\{\mathbf{M}^{I*}\}$ such that given the corresponding $\{M^{I*}, \bar{Q}^{I*}, \Pi^{I*}\}$ a period equilibrium exists with $M_t^{I*} = M_{t-1}^{I*}$, $\bar{Q}_t^{I*} = \bar{Q}_{t+1}^{I*}$, and $\Pi_t^{I*} = \Pi_{t-1}^{I*}$. Due to the $1:1$ correspondence between $\bar{Q}^{I*}$ and Π^{I*}, we sometimes omit Π^{I*} from the definition.

3 Benchmark Analysis

To establish a benchmark and to build intuition regarding our main results, we begin by analyzing the model without homophily, no discrimination and no quotas: that is, where $u_{q_i}(\cdot)$ does not depend on ν^I, $f(\cdot)$ depends on $\bar{Q}$ instead of $\bar{Q}^I$, and $d = 0$. Since identity is not directly payoff relevant in the this analysis, we focus on characterizing "identity neutral" symmetric equilibria where $\bar{Q}_t^A = \bar{Q}_t^B = \bar{Q}_t$ and $\sigma_{q,t}^A = \sigma_{q,t}^B = \sigma_{q,t}$ as the relevant benchmark.

Period Equilibria

Before characterizing the steady states, we detail period equilibria given average senior prestige and quality, and therefore omit the period notation, t, for the first part of our analysis.

We introduce the following notation for the expected utility of entry into the profession given strategies σ_{-i} are equal to σ_q:

$$U_{q_i}(\Pi, \bar{Q}, m_h) = E[u_{q_i}(\pi_i, \Pi) | \sigma_q], \tag{3}$$

Where $U_h(\Pi, \bar{Q}, m_h)$ can be represented as a function of m_h, since m_h is unique given σ_q, while $U_l(\Pi, \bar{Q})$ does not depend on m_h since $f(l, \bar{Q}^I, m_h) = 0$. This allows us to characterize the best response functions as follows:

$$\sigma_i = \begin{cases} 1 & \text{if } U_{q_i}(\Pi, \bar{Q}, m_h) - c > o_{q_i}, \\ \sigma \in [0,1] & \text{if } U_{q_i}(\Pi, \bar{Q}, m_h) - c = o_{q_i}, \\ 0 & \text{if } U_{q_i}(\Pi, \bar{Q}, m_h) - c < o_{q_i}. \end{cases}$$

That is, in equilibrium career entrants will apply to become juniors to the point where the expected utility from applying is equal to or less than the outside option.

Since all juniors are indifferent between applying and the outside option in any interior equilibrium this implies that the expected utility conditional on entry relative to the outside option must be the same for both high and low-type juniors. Formally:

Lemma 1. *In equilibrium, if* $\sigma_h, \sigma_l > 0$*, then the following condition must hold:*

$$U_l(\Pi, \bar{Q}) - o_l = U_h(\Pi, \bar{Q}, m_h) - o_h. \tag{4}$$

Lemma 1 illustrates the basic structure of period equilibria: Given $\bar{Q}$ and Π, the expected utility of low-type juniors is fixed. However, the expected utility for high types is decreasing in m_h due to the competition effect ($\partial f(h, \bar{Q}^I, m_h)/\partial m_h < 0$). Therefore, in equilibrium, high types will apply to training to the point where $U_h(\Pi, \bar{Q}, m_h)$ is low enough for (5) to hold.

Lemma 1 also shows that period equilibria can be characterized indirectly in terms of the equilibrium size of the sets of juniors, $\{m_h^*, m_l^*\}$, rather than referring to σ_q. Moreover, rearranging Condition 4,

$$U_h(\Pi, \bar{Q}, m_h) = U_l(\Pi, \bar{Q}) + o_h - o_l. \tag{5}$$

we see that an interior value of m_h^* is implicitly defined as a function of $\bar{Q}$.

Take m_h' to be the value of m_h that satisfies Condition (5)—this value is unique due to the monotonicity of the utility function in π_i. The following result establishes that, in terms of the equilibrium size of the sets of juniors, the period equilibrium is unique.

Corollary 1. *The period equilibrium,* m_h^*, m_l^**, is unique and* m_h^* *is characterized* *by:*

$$m_h^* = \begin{cases} 0 & \textit{if } m_h' \leq 0, \\ m_h' & \textit{if } m_h' \in (0, \lambda), \\ \lambda & \textit{if } m_h' \geq \lambda, \end{cases}$$

where m_h' is implicitly defined by (5).

Steady States

Next, we partially characterize the steady states of the model and establish that despite unique period equilibria, multiple steady states may exist. First we consider interior steady states and characterize the dynamics of quality, $\bar{Q}_t$—since we are considering an identity-neutral model with no discrimination, we can characterize a steady state by the average quality and prestige (recall that with $d = 0$, $\bar{Q}_t = \Pi_t$.), $\{\Pi, \bar{Q}\}$, rather than referring to set notations.

Note that $\bar{Q}_{t+1}$ is determined by the size of the set of juniors who realize non-zero quality (positive prestige) in period t, which is defined by the following equation:

$$|\{i : \pi_i > 0\}| = m_{h,t} f(h, \bar{Q}_t, m_{h,t}). \tag{6}$$

Since the set of juniors that realize positive senior quality in time t have π_i distributed uniformly over $[0, 1]$, the profession will hire all seniors who realize a prestige of $\pi_{L,t}$ or greater where $\pi_{L,t}$ satisfies the expression:

$$(1-\pi_{L,t})m_{h,t}f(h,\bar{Q}_t,m_{h,t})=1. \tag{7}$$

Moreover, since Π_{t+1} is characterized by the following expression:

$$\Pi_{t+1}=\frac{1-\pi_{L,t}}{2}, \tag{8}$$

we can substitute for $\pi_{L,t}$ using Equation 7 and use the fact that $\bar{Q}_{t+1}=\Pi_{t+1}$ to get the expression characterizing $\bar{Q}_{t+1}$ as a function of $\bar{Q}_t$:

$$\bar{Q}_{t+1}=[2m_{h,t}f(\bar{Q}_t,m_{h,t})]^{-1}. \tag{9}$$

That is, since Equation (5) implicitly characterizes $m_{h,t}$ as a function of $\bar{Q}_t$, Equation 9 characterizes the dynamics of quality, and can be used to identify interior steady states of the model.

However, when interior steady states of the model exist—i.e. $\bar{Q}^* \in (0,1)$ such that $\bar{Q}_{t+1}(\bar{Q}^*)=\bar{Q}^*$—they are not unique. As shown in the following proposition, a steady state of the model also exists at $\bar{Q}=0$.

Proposition 1. *$\bar{Q}^*=0$ is a steady state of the model. Interior steady states, $\bar{Q}^* \in (0,1)$, exist if and only if $\bar{Q}_{t+1}(\bar{Q}^*)=\bar{Q}^*$.*[7]

A corner solution with $\bar{Q}^*=0$ is a steady state since only low-talent juniors enter if $\bar{Q}_t=0$, and low-talent juniors do not realize positive quality, which implies

[7]Note that $\bar{Q}^*=1$ cannot be a steady state since the set of juniors that realize $Q_i=1$ has no mass.

that $\bar{Q}_{t+1} = 0$. The fact that there may be an interior and a corner steady state will be important when considering policy interventions since, as we show in the following section, with quotas and homophily it may be possible for the different identity groups to converge to different average senior quality, and in particular to a steady state where high-talent juniors of one identity type exit the profession.

Note that the steady states characterized above are "identity-neutral" in the sense that any composition of $\mathbf{M}^A$ and $\mathbf{M}^B$ constitute a steady state of the model as long as the corresponding $\bar{Q}$ is a steady state. Moreover, there is no persistence of identity at the steady states—if $M_t^A < M_t^B$ in period t, there exist $\mathbf{M}_{t+1}^A$ and $\mathbf{M}_{t+1}^B$ with $M_{t+1}^A = M_{t+1}^B$ that correspond to a period equilibrium as long as $\bar{Q}_t = \bar{Q}_{t+1}$, implying that a transition to equal representation can be achieved in a single period. As we show in the next section, however, this changes drastically when we introduce a preference for homophily to the model.

Persistence of Underrepresentation with Homophily Payoff

We first consider the case of no discrimination ($d = 0$). From a technical perspective the analysis of the model with a preference for homophily is similar to the analysis above, with the exception that juniors of a given identity type receive negative utility if they are over-represented relative to the mentoring capacity of seniors of their identity type.

Importantly, Lemma 1 continues to hold, and in each period equilibrium the expected relative value of entry is identical for all types that apply with positive probability.

Lemma 2. *All types that enters with positive probability in equilibrium ($\sigma_h^A, \sigma_l^A, \sigma_h^B$ or $\sigma_l^B > 0$) have an equal relative expected utility of entry, $U_{q_i}^I(\Pi^I, \bar{Q}^I, m_h, \nu^I) - o_{q_i}$.*

Next we link the analysis to the benchmark model without homophily by showing that for any steady state without homophily, there exists a corresponding steady state where representation, prestige and quality are all equal; i.e. $M^A = M^B$ and $\bar{Q}^A = \bar{Q}^B = \bar{Q}^*$.

Proposition 2. *If $\bar{Q}^*$ to be a steady state without homophily, then $\{M^A, M^B, \bar{Q}^A, \bar{Q}^B\}$ is a steady state if $\bar{Q}^A = \bar{Q}^B = \bar{Q}^*$.*

Intuitively, we can think of the steady state with equal quality in both groups and homophily as two separate professions for $I = A, B$. If both professions are at a steady state, i.e. if $\bar{Q}^A = \bar{Q}^B = \bar{Q}^*$, then the overall profession is at a steady state as well, and juniors will apply in proportion to the number of seniors of their identity type. Encouragingly, Proposition 2 shows that homophily does not need to have a distortionary impact on the profession: no matter how strong the preference for homophily is, a steady state exists with equal representation and equal quality.

While this may give the impression that homophily will not distort the profession, note that Proposition 2 shows that unequal representation is also a steady state of the model with homophily. Importantly, however, transitioning to equal representation is no longer possible in equilibrium. That is, as shown in the following corollary, if the profession is at a steady state with $\bar{Q}^A = \bar{Q}^B$ and type A is underrepresented ($M_t^A < M_t^B$), this underrepresentation will persist in all future periods since $M^A_{t+1} = M^A_t$.

Corollary 2 (Persistence of under-representation)**.** *If the profession is at a steady state with $\bar{Q}^A = \bar{Q}^B > 0$, then $M^I_t = M^I_{t+1}$ in all period equilibria.*[8]

Corollary 2 shows that given a preference for homophily, if the profession starts at a steady state with $M^A < M^B$, that imbalance will persist in perpetuity. The intuition for the persistence of under-representation is the fact that for representation of type A to increase in period $t+1$, it must be the case that a higher proportion of high-talent juniors of type A entered in period t relative to $t-1$. However, this cannot be a period equilibrium since it implies that expected value of entry is lower for high-talent juniors of type A than for high-talent juniors of type B, $U^A_h(\Pi^A, \bar{Q}^A, m_h, \nu^A) < U^B_h(\Pi^B, \bar{Q}^B, m_h, \nu^B)$, which violates the condition for a period equilibrium in Lemma 2.

Importantly, Corollary 2 shows that if one of the two types is underrepresented, then underrepresentation will persist. Therefore, to transition to equal representation, a policy intervention will be necessary.

Discrimination: Lastly, we consider the impact of discrimination against type A ($d > 0$) on the steady states of the model with homophily. First, note that corner steady states ($\bar{Q} = 0$) are unaffected. For interior values, however, $d > 0$ implies that the average quality of seniors of type A is higher than for type B. This is due to the fact that the market hires all seniors with prestige higher than some value,

[8] This corollary follows directly from Lemma 4, which is presented in the proof of Proposition 2.

π_L. Therefore, $\Pi^A = \Pi^B$ in all period equilibria, which will result in $\bar{Q}^A > \bar{Q}^B$ due to discrimination.

On one hand, $\bar{Q}^A > \bar{Q}^B$ means that the probability that high-talent juniors of type A realize strictly positive quality/prestige is higher than that of a high-talent junior of type B. On the other hand, the average prestige of type A juniors is lower due to discrimination. To be a steady state, it must be the case that the effect of discrimination on type A is balanced by the higher average quality of mentors of type A, in the sense that despite the discrimination the probability that a high-skilled junior realizes $\pi_i \geq \pi_L$ is the same for type A and type B. Formally:

Proposition 3. *At a steady state with* $d > 0$, $\{M^{A*}, M^{B*}, \Pi^*, \bar{Q}^{A*}, \bar{Q}^{B*}\}$*, and* $\Pi^* > 0$ *the following condition is satisfied:*

$$\Pr(\pi_i > \pi_L^* | q_i = h, I = A, \bar{Q}^{A*}, m_h^*) = \Pr(\pi_i > \pi_L^* | q_i = h, I = B, \bar{Q}^{B*}, m_h^*) \qquad (10)$$

Intuitively, suppose that Condition 10 is violated and that in period t the probability that a high type of identity A realizes $\pi_i > \pi_L^*$ is higher than for a high type of identity A. However, as shown above, the ratio of high-talent juniors to seniors is the same for both types. Therefore, a higher probability of realizing $\pi_i > \pi_L^*$ implies that the number of seniors of type A will increase in period $t+1$ ($M_{t+1}^A > M_t^A$), which shows that $\{M^{A*}, M^{B*}, \Pi^*, \bar{Q}^{A*}, \bar{Q}^{B*}\}$ is not a steady state.

Proposition 3 also speaks to the difficultly of observing discrimination if a profession is in a steady state. Despite the fact that $\bar{Q}^{A*} > \bar{Q}^{B*}$, this imbalance is endogenously "hidden" since discrimination against juniors of type A is perfectly offset by the higher average quality of the seniors of type A. Therefore, the dynamics of the observable metric of prestige is the same with and without discrimination: despite discrimination, an equal proportion of juniors of both identity types are hired as seniors and seniors of both identity types have the same average prestige.

4 Achieving Equal Representation

In this section we will analyze the effectiveness of quotas as a policy tool to transition to a steady state with equal representation and no stigma. We define stigma as the case where $\Pi_t^A < \Pi_t^B$; note that, due to discrimination, stigma does not imply that there is a quality difference between seniors of different identity types. We do not consider an explicit welfare objective or explicitly model the benefits of equal representation. However, beyond fairness concerns, we emphasize that the literature has highlighted many potential benefits of equal representation in high-skill professions (see Auriol et al., 2022).

We focus on a setting with discrimination $d > 0$—we discuss the impact of eliminating discrimination in Subsection 4.3—and transitions from an initial point of equal prestige and quality ($\Pi^A = \Pi^B$, $\bar{Q}^A = \bar{Q}^B$) that corresponds to a steady state of the model with no homophily, $\bar{Q}^* > 0$, that is asymptotically stable. We discuss the stability of steady states in the appendix and show that, generically, at least one interior steady state is stable. Moreover, this assumption is not restrictive since there always exists a transition path from an unstable steady state to a stable steady state.

4.1 Analysis of quotas on underrepresented juniors

A natural starting point is considering a quota on juniors. In particular, we consider the effectiveness of instituting a 1:1 quota on juniors as a policy for achieving a transition to equal representation at the senior level. We show that while such a policy will mechanically equalize the number of juniors of each identity type, it will be surprisingly ineffective when it comes to equalizing representation at the *senior level*. The reason a quota on juniors is ineffective is due the familiar problem of selection—simply put, a quota on juniors will not necessarily increase the proportion of high-talent juniors of the under-represented type.

Formally, we model a 1:1 quota as two separate entry lotteries for types A and B, where all applicants of type A are randomly selected to fill $\lambda/2$ junior slots, and all applicants of type B are randomly selected to fill the remaining $\lambda/2$ slots. Surprisingly, as illustrated in the following result, there is no impact of junior quotas on representation in most cases.

Proposition 4 (Ineffectiveness of junior quota)**.** *If the profession is at a steady state* $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ *with* $\Pi^{A*} = \Pi^{B*} = \Pi^* > 1/2$ *and* $m_h^{B*} \leq \lambda/2$*, then* $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ *remains a steady state under a 1:1 quota on juniors.*

The intuition for this result is as follows: A quota on juniors of type A in period t does not increase the relative attractiveness of entry for the high-talent juniors of type A since M_t^A and Π^A remain unchanged. Since the number of high-talent juniors that enter depends only on the relative utility of entry (rather than the probability of entry), a 1:1 quota on juniors will not result in more entry of *high-talent* juniors of the underrepresented type. That is, if type A is underrepresented ($M^A < M^B$) at the initial steady state, then the junior quota will be filled by low-talent juniors of type A, who are not hired by the profession as seniors and therefore do not impact representation at the senior level.[9]

4.2 Analysis of quotas on underrepresented seniors

Here we consider a minimum quota on seniors of the underrepresented type as a method for achieving equal representation. We first consider the implementation of a 1:1 quota in period t and higher; i.e. the profession is constrained to hire $M^A_{t+1} = M^B_{t+1} = 1/2$. Note that a quota on seniors does not directly impact the entry decisions of the juniors, since juniors receive utility based on their realized senior quality regardless of whether they are hired by the profession. However, as discussed above, a quota will affect entry decisions indirectly through its impact on prestige—i.e. stigma. Also, note that since we are considering the case where $d > 0$, quotas do not necessarily result in a situation where average quality is lower among seniors of type A.

Note that if the profession starts from a point of underrepresentation, $M^A < M^B$, the implementation of a 1:1 quota in period t will result in stigma. Specifically, a binding quota will require different prestige cutoffs for seniors of different

[9] A quota on juniors can only have an impact on the profession if (1) it decreases the number of high-talent juniors of the over-represented type, which only happens if $m^B_h > \lambda/2$, or (2) if the profession hires seniors with $\pi_i = 0$, which is only the case if $\Pi_t < 1/2$.

identity types ($\pi^A_{L,t} < \pi^B_{L,t}$). This implies that the average prestige of the underrepresented seniors in period $t+1$ will be lower than for the overrepresented type ($\Pi^A_{t+1} < \Pi^B_{t+1}$), which will have an impact of the entry decision of career entrants in period $t+1$. Accordingly, as shown in the following lemma, a 1:1 quota on seniors will not result in 1:1 entry of juniors due to the perceived difference in quality of seniors of different identity types.

Lemma 3 (Crowding Out)**.** *If $\Pi^A_t < \Pi^B_t$, then $\nu^B_t > 0$ in the period equilibrium.*

That is, juniors of the identity-group with higher average prestige will enter in a higher proportion relative to the proportion of seniors of that identity group—there is a "crowding out" effect of stigma. In fact, if the impact of the quota on prestige is high enough, then the crowding out effect can cause the profession can converge to an asymmetric steady state where high-talent juniors of the underrepresented type select out of the profession.

Proposition 5 (Adverse selection)**.** *If the profession is at a steady state $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ with $\Pi^{A*} = \Pi^{B*} = \Pi^* > 0$, then for all $M^{A*} < M^{A'}$ for some $M^{A'} > 0$, a 1:1 quota on seniors will result in a transition to $\bar{Q}^A = 0$.*

Proposition 5 follows from the fact that with homophily, the equilibrium entry decisions of high-talent juniors are driven by the average prestige of seniors in their identity-group. In particular, given Lemma 3 we can show that if stigma is high enough, then the profession will transition to a steady state where all high-talent juniors of the underrepresented type select out of the profession. This effect is then permanent since if all high talent juniors of type A select out the profession in a given period due to stigma, then the average quality of seniors of type A will equal 0 in the following period, causing all high talent juniors of type A select out the profession in the following period (and all future periods) as well. This shows that the stigma introduced by an abrupt quota can be self-fulfilling, in the sense that it causes a transition to a steady state with permanent stigma and lower quality for the underrepresented type.

Encouragingly, however, our next results shows that a transition to equal representation and equal prestige is always feasible as long as the transition is gradual enough. For the purpose of illustration, we first prove the result for the case of discontinuity of utility at $\nu = 0$, in the sense that limit of $U^I_{q_i}(\Pi^I, \bar{Q}^I, m_h, \nu^I) - U^I_{q_i}(\Pi^I, \bar{Q}^I, m_h, 0)$ as ν approaches zero from above is equal to some constant strictly greater than zero.

Proposition 6 (Gradual Transition)**.** *If* $\lim_{\nu \to 0^+}[U^I_{q_i}(\Pi^I, \bar{Q}^I, m_h, \nu^I) - U^I_{q_i}(\Pi^I, \bar{Q}^I, m_h, 0)] = r > 0$ *and the profession is at a steady state* $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ *with* $\Pi^{A*} = \Pi^{B*} = \Pi^* > 0$ *and* $M^{A*} < M^{B*}$*, then there exists a monotonically increasing sequence of quotas on seniors of type A,* $\{\bar{M}_t\}$*, that results in a convergence to a steady state with* $\Pi^A = \Pi^B = \Pi^*$ *and* $M^A = M^B$*.*

Proposition 6 shows that when there is a discontinuity at $\nu = 0$, which implies that for small differences in Π^A and Π^B, the equilibrium identity ratio of the juniors will equal the identity ratio of the seniors—that is, in equilibrium $\nu = 0$ for small differences Π^A and Π^B—then a transition is possible as long it is gradual. Note that since the dynamics of quality and prestige are independent of the size of the profession, and a small increase in $\{\bar{M}_t\}$ in each period corresponds to a small decrease in the relative prestige of seniors of type A. Moreover, given that Π^* is a steady-state in the "A-profession," small deviations in Π^A imply that the dynamics of the model point back to Π^*. Therefore, a gradual increase in the quota, and correspondingly small increases $\{\bar{M}_t\}$, ensures that Π_t^A stays within the "basin of attraction" of Π^* along the whole path of transition. This implies that if the increase in the quota is gradual enough, then the profession will transition to a stigma-free steady state with equal representation and avoid the pitfall of adverse selection highlighted in Proposition 5.

Given the result of Proposition 6, it might be natural to assume that the same dynamic path of quotas will also result in a transition to equal representation without a discontinuity at $\nu = 0$ since continuity implies that the impact homophily approaches 0 as ν approaches 0. This, however, may not be the case due to the crowding out effect highlighted in Lemma 3. That is, the crowding out effect of a quota could lead to a dynamic where instead of equalizing, Π_t^A and Π_t^B grow farther apart. However, there is a simple remedy: by instituting a quota on seniors *and* a corresponding quota on juniors, the crowding out effect can be directly eliminated.

Corollary 3 (Coordinated Quotas)**.** *If the profession is at a steady state* $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ *with* $\Pi^{A*} = \Pi^{B*} = \Pi^* > 0$ *and* $M^{A*} < M^{B*}$*, then there exists a monotonically increasing sequence of quotas,* $\{\bar{M}_t^A, \bar{m}_t^A\}$*, that results in a convergence to a steady state with* $\Pi^A = \Pi^B = \Pi^*$ *and* $M^A = M^B$*.*

The intuition for Corollary 3 is straightforward. By constraining the set of juniors to be proportional to the set of seniors (in terms of identity-types), quotas on juniors eliminate the crowding out effect of perceived differences in senior quality—essentially, the quota establishes separate professions for the two different identity types, which means that the result of Proposition 6 applies and a transition to a stigma-free steady state with equal representation can be achieved.

4.3 Discussion of proposed policies

Lastly, we discuss the implications of our analysis for measures that have been proposed or implemented for addressing underrepresentation.

Eliminating discrimination: First, we consider the possibility of simply eliminating discrimination. While it is unclear whether such a policy is feasible in our setting, eliminating discrimination is of course a crucial objective and has social value on its own. Moreover, the following result shows that eliminating discrimination would impact underrepresentation.

Corollary 4 (Discrimination and underrepresentation)**.** *Take $d > 0$. If the profession is at a steady state $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ with $\Pi^{A*} = \Pi^{B*} = \Pi^* > 0$ and discrimination is eliminated in period t ($d = 0$ in period t), then $M^A_{t+1} > M^A_t$.*

Corollary 4 follows from the fact that if discrimination is eliminated in period t, then the average prestige in of seniors of type A will increase in period t and more juniors of type A will enter, causing $M^A_{t+1} > M^A_t$.

While Corollary 4 shows that there is an impact of eliminating discrimination on underrepresentation, it will not necessary lead to *equal representation*: while

$M^A_{t+1} > M^A_t$, it is not true that $M^A_{t+1} = 1/2$ in all cases. Moreover, in period $t+1$ both prestige and quality will be equal for both groups of seniors, implying that if $M^A_{t+1} < 1/2$, then underrepresentation will be persistent even absent discrimination. This implies that rather than a substitute, policies to eliminate discrimination should be seen as a complement to quotas to address underrepresentation.

The cascade model: Many professions that exhibit underrepresentation at the senior level suffer from the so-called "leaky pipeline," where the level of representation is high at junior levels, but decreases in seniority (see Buckles, 2019). One example of this is academia, where even in fields that are close to parity at the undergraduate level, women are increasingly underrepresented at the level of PhD, Assistant, Associate and Full professor. The cascade model, used in Sweden and Germany (Wallon et al., 2015), is meant to address the leaky pipeline by setting a soft quota at each level of seniority that is equal to the level of representation at the level below.

In the context of our model, it is natural to interpret bachelor's students as career entrants, PhD students as juniors, and professor positions as seniors. In accordance with the leaky pipeline, consider a field that features equal representation at the bachelor's level, some underrepresentation at the PhD level, and higher underrepresentation at the professor level. Applied strictly, the cascade model would transition to equal representation at the professor level in two periods. Somewhat surprisingly given the result of Proposition 6, we find that the cascade model can result in a *higher stigma relative to instituting a 1:1 quota in the first period.*

To explain the intuition behind this result, note that under the cascade model the initial quota at the professor level will be binding and therefore the prestige of the seniors of the underrepresented type will lower in $t+1$ relative to the steady state. This lowers the expected value of entry into the PhD program for bachelor's students of the underrepresented type and despite the binding 1:1 quota on underrepresented PhD students, can result in a decrease to the *total* number of underrepresented high-talent PhD students in $t+1$ relative to t. In this case, the total number of underrepresented seniors that realize high quality in period $t+1$ is lower than in period t. Therefore, instituting a 1:1 quota in period $t+1$ will require a lower $\pi^A_{L,t+1}$ relative to the $\pi^A_{L,t}$ that would have been required in the case where a 1:1 quota is instituted in period t. This shows that the cascade model can result in a lower perceived quality relative to jumping straight to a 1:1 quota.

We present this result formally in the following proposition (the result extends straightforwardly to cascade quotas that take more than two periods to reach a 1:1 quota). Take $\bar{M}$ equal to a 1:1 quota in period t, and $\bar{M}'$ equal to a quota where $\bar{M}'_{t+1}$ is equal to a 1:1 quota and $\bar{M}'_t < \bar{M}'_{t+1}$ is binding.

Proposition 7. *If the profession is at a steady state* $\{\mathbf{M}^{A*}, \mathbf{M}^{B*}\}$ *with* $\Pi^{A*} = \Pi^{B*} = \Pi^* > 0$ *and* $M^{A*} < M^{B*}$. *If* $U^A_l(\Pi^*, \bar{Q}^*, m_h, \nu^A_t) - U^A_l(\Pi^*, \bar{Q}^*, m_h, \nu^A_{t+1}) < U^A_h(\Pi^*, \bar{Q}^*, m_h, \nu^A_t) - U^A_h(\Pi^*, \bar{Q}^*, m_h, \nu^A_{t+1})$ *under* $\bar{M}'$, *then* Π^A_{t+2} *under* $\bar{M}'$ *is* lower *than* Π^A_{t+1} *under* $\bar{M}$.

Proposition 7 provides more detail about the sequence of quotas that will transition to equal representation and quality (Proposition 6): instead of a continuous increase, it may be beneficial to hold the quota constant at an intermediate level for a number of periods to allow the prestige of seniors of the underrepresented type to move closer to Π^*.

Tie-breaker models: A common measure used to address underrepresentation is to favor underrepresented candidates in cases of "equal quality," which we refer to as a tie-breaker quota. In the context of our model, where quality is not directly observable, a tie-breaker quota would entail a preference for underrepresented candidates in cases of equal prestige. Moreover, since we consider at a continuous prestige/quality distribution this quota would be non-binding. However, our analysis extends naturally to a case where the observed signal of senior quality is coarse. In this case, favoring underrepresented seniors of equal prestige will result in stigma since the quota implies that all seniors with the marginal level of prestige are underrepresented seniors.

On one hand, a tie-breaker quota avoids the problem illustrated in Proposition 7 by construction, since the quota endogenously limits the difference in prestige between seniors of the two identity categories. That is, as long as the signal of quality is not too coarse, then both Π_t^A and Π_t^B will stay within the catchment area of Π^*. On the other hand, we cannot exclude the possibility that a tie-breaker quota will lead to cycling rather than a transition to equal representation.

That is, while implementing a tie-breaker quota in period t will lead to an increase in M_{t+1}^A, the impact on M_{t+2}^A, relative to M_t^A, is unclear due to the problem of crowding out. Since the quota implies that $\Pi_{t+1}^A < \Pi_{t+1}^B$, we know that $M_{t+2}^A < M_{t+1}^A$ by Lemma 3. Therefore, we cannot exclude the possibility that $M_{t+2}^A \leq M_t^A$

which could lead to a cycle about the original levels M^{A*}, M^{B*}. Therefore, it may be possible that a tie-breaker quota could require an occasional "nudge"—a discrete increase in the number of underrepresented seniors—to put it on the path to transition to equal representation.

5 Conclusion

In this paper we explore the trade-off between representation and perceived quality, and establish a dynamic argument for quotas to correct for underrepresentation even if they introduce stigma in the short run. Our research provides also provides important insights into policy measures. First, it is not sufficient to simply eliminate discrimination, or to institute a quota at the junior level. Instead, a quota at the senior level is necessary to achieve equal representation. Second, due to adverse selection, an abrupt transition to equal representation can cause permanent stigma and real quality difference between seniors of the two identity types, while a gradual transition can result in a stigma-free steady state with equal representation. That is, while transition may require some stigma in the profession in the short run, this stigma is temporary as the profession transitions to a long-term stigma-free steady state with equal representation.

Moreover, we show that a "cascade model," where employment at the senior level is equalized to the identity proportions at the junior level, can be counterproductive relative to a quick transition to equal representation. Instead, a preference for underrepresented seniors in the case of equal quality seems preferable, although it may require additional nudges to transition all the way to equal representation.

Our research also suggests that an important avenue for future research on role models is to explore the interaction of identity and quality. That is, while the empirical literature has focused on characterizing the impact of the identity of role models on educational and career choices, the impact of the quality of role models on choices is largely unexplored. Our research suggests that while more role models of an underrepresented type may be an important factor in combating underrepresentation, such a strategy could backfire even if role models of the underrepresented type are *perceived* to be of lower quality by potential entrants. Therefore, our research highlights that empirical evidence on the interaction of identity-quotas, career choice and perceived quality is essential when it comes to addressing the effectiveness of policy to achieve equal representation.

References

Athey, S., Avery, C., and Zemsky, P. (2000). Mentoring and diversity. *American Economic Review*, 90(4):765–786.

Auriol, E., Friebel, G., Weinberger, A., and Wilhelm, S. (2022). Underrepresentation of women in the economics profession more pronounced in the united states compared to heterogeneous europe. *Proceedings of the National Academy of Sciences*, 119(16):e2118853119.

Battaglini, M., Harris, J., and Patacchini, E. (2022). Professional interactions and hiring decisions: Evidence from the federal judiciary. *Journal of Labor Economics*, Forthcoming.

Besley, T., Folke, O., Persson, T., and Rickne, J. (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review*, 107(8):2204–42.

Buckles, K. (2019). Fixing the leaky pipeline: Strategies for making economics work for women at every stage. *Journal of economic perspectives*, 33(1):43–60.

Carvalho, J.-P. and Pradelski, B. (2022). Identity and underrepresentation: Interactions between race and gender. *Journal of Public Economics*, 216:104764.

Coate, S. and Loury, G. C. (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5):1220–1240.

European Commission (2019). She figures 2018. Technical report, Directorate-General for Research and Innovation.

Fang, H. and Moro, A. (2011). Chapter 5 - theories of statistical discrimination and affirmative action: A survey. volume 1 of *Handbook of Social Economics*, pages 133–200. North-Holland.

Fershtman, D. and Pavan, A. (2021). “soft” affirmative action and minority recruitment. *American Economic Review: Insights*, 3(1):1–18.

Gaule, P. and Piacentini, M. (2018). An advisor like me? advisor gender and post-graduate careers in science. *Research Policy*, 47(4).

Heilman, M., Block, C., and Lucas, J. (1992). Presumed incompetent? stigmatization and affirmative action efforts. *Journal of Applied Psychology*, 77(4):536–544.

Hilmer, C. and Hilmer, M. (2007). Women helping women, men helping women? same-gender mentoring, initial job placements, and early career publishing success for economics phds. *American Economic Review*, 97(2):422–426.

Leslie, Lisa, M. D. K. D. (2014). The stigma of affirmative action: A stereotyping-based theory and meta-analytic test of the consequences for performance. *The Academy of Management Journal*, 57(4):964–989.

Müller-Itten, M. and Öry, A. (2022). Mentoring and the dynamics of affirmative action. *American Economic Journal: Economic Policy*, 14(2):402–44.

Porter, C. and Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–54.

Riise, J., Willage, B., and Willén, A. (2022). Can Female Doctors Cure the Gender STEMM Gap? Evidence from Exogenously Assigned General Practitioners. *The Review of Economics and Statistics*, 104(4):621–635.

Siniscalchi, M. and Veronesi, P. (2020). Self-image bias and lost talent. Working Paper 28308, National Bureau of Economic Research.

Wallon, G., Bendiscioli, S., and Garfinkel, M. (2015). Exploring quotas in academia. Technical Report 12.5.8050.0239.0, Robert Bosch Stiftung.

6 Appendix: Proofs

Proof Lemma 1: The best response functions specify that for interior equilibria ($\sigma_l, \sigma_h > 0$):

$$E[u_i|\hat{a}_i = 1, \bar{Q}, q_i, \sigma_{-i}] = o_{q_i}.$$

Let p be the the probability of entry given σ_l, σ_h. This allows us to rewrite the condition for an interior equilibrium as a function of the expected wage conditional on entry:

$$pE[w_i|a_i = 1, \bar{Q}, m_h, q_i] - (1-p)o_{q_i} - c = o_{q_i},$$

which simplifies to:

$$E[w_i|a_i = 1, \bar{Q}, m_h, q_i] - o_{q_i} = \frac{c}{p}.$$

Since the right-hand side of the expression above is the same for high and low-talent juniors, the left-hand side must be equal for $q_i = l, h$ in an interior equilibrium.

$$E[w_i|a_i = 1, \bar{Q}, m_h, l] - o_l = E[w_i|a_i = 1, \bar{Q}, m_h, h] - o_h. \tag{11}$$

Before proceeding with the proof, we show the following corollary:

Corollary 5. *$m_h \geq \lambda M_H$ in all equilibria.*

Corollary 5 follows from (**??**) and the assumption that $\beta - o_l < 1/2 - o_h$: if $m_h < \lambda M_H$, then all high type juniors who enter will be matched to the highest-quality group and realize an expected payoff of $1/2 - o_h$. That is, the expected utility conditional on entry for the high type is greater than for the low type ($\beta\bar{Q} - o_l \leq \beta - o_l$), which violates (**??**). Therefore, in equilibrium, m_h must be greater or equal to λM_H.

This implies that no low-talent juniors will match with the group of highest quality mentors and the expected wage for a low type is equal to:

$$E[w_i|\bar{Q},l] = \beta\bar{Q}, \tag{12}$$

which gives the equilibrium condition:

$$\beta\bar{Q} - o_l = g(\bar{Q}, m_h) - o_h.$$

□

Proof of Proposition 1: In the main text we establish that an interior period equilibrium exists if and only if $m_h' \in (0, \lambda)$ where m_h' is defined in (**??**). We complete the proof by showing that a corner equilibrium exists with $m_h^* = 0$ if and only if $m_h' \leq 0$, and with $m_h^* = \lambda$ iff $m_h' \geq 0$.

First, take $m_h' \leq 0$ and assume an equilibrium exists with $m_h^* > 0$. Note that $m_h' \leq 0$ implies that:

$$\beta\bar{Q} - o_l \geq g(\bar{Q}, 0) - o_h.$$

Since $g(\bar{Q}, 0)$ is strictly decreasing in m_h, this equation shows that at m_h^*, $\beta\bar{Q} - o_l < g(\bar{Q}, m_h^*) - o_h$. Therefore, m_h^* cannot be an equilibrium. However, since

$\beta\bar{Q} - o_l > 0$ for all $\bar{Q}$, $m_h^* = 0$, $m_l^* = \lambda$ is an equilibrium.

The proof for $m_h' \geq 0$ is analogous. Since $\beta\bar{Q} - o_l \leq g(\bar{Q}, 1) - o_h$, this expression also holds for all $m_h < \lambda$, implying that $m_h^* = \lambda$, $m_l^* = 0$ is the unique equilibrium. □

Proof of Proposition 1: First, note that if $\bar{Q}_t = 0$, then it is a best response for all high-type juniors to set $\hat{a}_i = 0$ since the relative expected utility of applying is negative: i.e. $o_h > 1/2f(0) - c$. Since $o_l + c < 0$, however, the relative expected utility of applying is positive for the low type if the probability of entry is equal to one. Therefore, $m_{h,t} = 0$ and $m_{l,t} = \lambda$ in any period equilibrium, and $\bar{Q}_{t+1} = 0$.

Second, trivially, $\bar{Q}^*$ is not a steady state if $\bar{Q}_{t+1}(\bar{Q}^*) \neq \bar{Q}^*$. If $\bar{Q}_{t+1}(\bar{Q}^*) = \bar{Q}^*$, however, then $\bar{Q}^*$ is a steady state since each $\bar{Q}$ is associated with a unique distribution of Q_j (uniform between $[Q_L, 1]$). □

Proof of Lemma 2: The proof follows directly from the proof of Lemma 1.

Proof of Proposition 2: First, note that the result is trivial for $\bar{Q} = 0$ since given $\bar{Q}_t = 0$, only low-talent juniors will apply and in equilibrium they will apply in proportion to M_t^A and M_t^B. Therefore, $\bar{Q}_{t+1} = 0$ and $M_t^A = M_{t+1}^A$.

For $\bar{Q}^A = \bar{Q}^B > 0$, we first introduce the following result characterizing period equilibria:

Lemma 4. *If $\bar{Q}^A = \bar{Q}^B > 0$, then in equilibrium:*

$$\frac{m_h^A}{m_h^B} = \frac{m_l^A}{m_l^B} = \frac{M^A}{M^B}.$$

That is, juniors will enter in the same identity-proportion as the proportion of mentors.

This lemma follows from Lemma 2: given the equal quality in both identity groups, it cannot be the case that any juniors face a positive probability of matching with an out-group mentor, since that would imply a lower expected utility of entry for that group. Moreover, $\frac{m_h^A}{m_h^B} = \frac{m_l^A}{m_l^B}$, since otherwise the high-type of one identity group would have a higher probability of matching with a high-quality mentor. ■

Lemma 4 shows that given $\bar{Q}_t^A = \bar{Q}_t^B$, $\Pr(I_{\neg i}|I_i) = 0$ in equilibrium for both identity types. This implies that the following equation defines an interior solution for m_h^{I*}:

$$m_h^{I'} = \lambda M_H^I \left[\frac{1 - f(\bar{Q})}{(\beta\bar{Q} - o_l) - (f(\bar{Q}) - o_h)} \right], \tag{13}$$

That is, $\frac{m_{h,t}^A}{m_{h,t}^B} = \frac{M_{H,t}^A}{M_{H,t}^B}$. In turn, this shows that if $\bar{Q}_t^A = \bar{Q}_t^B = Q^*$, where Q^* corresponds to a steady state, then $\bar{Q}_{t+1}^A = \bar{Q}_{t+1}^B = Q^*$, since $m_{h,t}^A$ and $m_{h,t}^B$ are both proportional to the size of the sets of mentors, M_t^A and M_t^B. □

Next we briefly address the existence of a stable interior steady state. Visually, note that interior steady states exist at points where $\bar{Q}_{t+1}(\bar{Q})$ either cross or are tangent to the 45 degree line. Moreover, $\bar{Q}_{t+1}(\bar{Q})$ is below the 45 degree line at both endpoints, 0 and 1, by assumption. It follows by the continuity of $\bar{Q}_{t+1}(\bar{Q})$, that as long as the function crosses the 45 degree line at some interior point, a steady state exists that where $\bar{Q}_{t+1}(\bar{Q})$ crosses the 45 degree line from above.

This gives the following result which completes the proof:

Result 1. *If $\bar{Q}_{t+1}(\bar{Q}) > \bar{Q}$ for some $\bar{Q} \in (0,1)$, then a stable steady state exists.*

Proof: Note that by the argument above, $\bar{Q}_{t+1}(\bar{Q}) > \bar{Q}$ for some $\bar{Q} \in (0,1)$ implies a steady state exists that where $\bar{Q}_{t+1}(\bar{Q})$ crosses the 45 degree line from above at some $\bar{Q}'$. This implies that the linearization of $\bar{Q}_{t+1}(\bar{Q})$ at $\bar{Q}'$ has a strictly negative slope, which implies that the eigenvalue criterion for a stable steady state is satisfied. □

Proof of Proposition 4: First, note that if $M^{A*} = M^{B*}$ then the 1:1 quota is redundant. Therefore, assume without loss of generality that $M^{A*} < M^{B*}$ and that the quota is initiated in period t. Therefore, a quota implies that with some probability each junior of type A will match with a mentor of type B.

$$\Pr(A|B) = \frac{\lambda/2 - \lambda M^A}{\lambda/2}$$

The expected utility of type A juniors conditional on entry in period t is:

$$(1 - \Pr(A|B))E[w_i|\bar{Q}_t^A, q_i] + \Pr(A|B)(E[w_i|\bar{Q}_t^B, q_i] - \eta).$$

However, since $\bar{Q}_t^A = \bar{Q}_t^B = \bar{Q}^*$ in period t, this expression simplifies to:

$$E[w_i|\bar{Q}^*, q_i] - \Pr(A|B)\eta.$$

That is, relative to no quota, in equilibrium the expected utility conditional on entry is lower by $\Pr(A|B)\eta$ for *both* the high and low type.

Therefore, $\Pr(A|B)\eta$ cancels out of the equilibrium condition listed in Lemma 2, and $m^A_{h,t}$ is still characterized by Equation 13 (in the proof of Proposition 2 above), which shows that the equilibrium level of $m^A_{h,t}$ is unchanged by the quota.

Next, note that $m^B_{h,t}$ is also unchanged by the quota by the same argument, and by the fact that $m^{B*}_h \leq \lambda/2$ (i.e. the size of the set of high-talent juniors of type B is lower than the quota at the steady state).

Since $m^A_{h,t} = m^{A*}_h$ and $m^B_{h,t} = m^{B*}_h$ the quota does not impact the set of juniors that realize positive quality in period t. Lastly, since $\bar{Q}^* \geq 1/2$, only mentors with strictly positive quality are hired by the profession ($Q_L > 0$). And since the set of juniors that realize positive quality in period t (with the quota) is identical to $t-1$ (without the quota), $Q_{L,t}$ will also be unchanged, and $\mathbf{M}^A_t = \mathbf{M}^{A*}$ and $\mathbf{M}^B_t = \mathbf{M}^{B*}$.
□

Proof of Lemma 3: First take the case of $m_l = 0$ and assume $\Pr(A|B) = 0$. In this case, both high-talent types must enter, and the following equation holds by Lemma 2:

$$\Pr(B|A)\left(g(\bar{Q}^B, m_h) - \eta\right) + (1 - \Pr(B|A))\left(g(\bar{Q}^A, m_h)\right) = g(\bar{Q}^B, m_h).$$

Since $m_l = 0$ and only high-talent juniors enter, $g(\bar{Q}^B, m_h) > g(\bar{Q}^A, m_h)$ which implies that $\Pr(A|B)$ must be strictly greater than zero for the above equation to hold.

Since $\Pr(A|B) > 0$, we can use the equilibrium condition in Lemma 2 to get the following expression for $\Pr(A|B)$:

$$\Pr(A|B) = \frac{\beta(\bar{Q}^B - \bar{Q}^A)}{\beta(\bar{Q}^B - \bar{Q}^A) + \eta},$$

Next, take the case of $m_l > 0$ and $\Pr(A|B) = 0$. First, consider $\beta\bar{Q}^B - \eta < \beta\bar{Q}^A$. In this case, both low-talent types enter if $m_l > 0$, and each type prefers to be matched to an own-type mentor. Therefore, the following equilibrium condition must hold by Lemma 2:

$$\Pr(B|A)\left(\beta\bar{Q}^B - \eta\right) + (1 - \Pr(B|A))\left(\beta\bar{Q}^A\right) = \beta\bar{Q}^B.$$

This is a contradiction since $\bar{Q}^A < \bar{Q}^B$, showing that $\Pr(A|B)$ must be strictly positive.

Solving for $\Pr(A|B)$ as above gives:

$$\Pr(A|B) = \frac{g(\bar{Q}^B, m_h) - g(\bar{Q}^A, m_h)}{g(\bar{Q}^B, m_h) - g(\bar{Q}^A, m_h) + \eta},$$

which is strictly greater than 0.

Lastly, assume that $\beta\bar{Q}^B - \eta \geq \beta\bar{Q}^A$, $m_l > 0$ and $\Pr(A|B) = 0$. In this case, low-types of both identities prefer to be matched with a B mentor, and therefore conditional on entry have the same probability of being matched with a B mentor. However, this implies that the above equilibrium condition cannot hold for low-talent junior, since the relative expected utility of entry is strictly higher for B-type juniors, which implies that $m_l^A = 0$.

This, however, implies that the following equilibrium condition cannot hold:

$$\Pr(B|A)\left(g(\bar{Q}^B, m_h) - \eta\right) + (1 - \Pr(B|A))\left(g(\bar{Q}^A, m_h)\right) \geq g(\bar{Q}^B, m_h).$$

Since $g(\bar{Q}^B, m_h) > g(\bar{Q}^A, m_h)$ given that $\bar{Q}^B > \bar{Q}^A$, and only high-talent juniors match with mentors of type A.

$\square$

Proof of Proposition 5: By Lemma 3, if $\bar{Q}_t^A < \bar{Q}_t^B$, then $P(A|B)_t > 0$ in equilibrium. This implies that a proportion of high-talent juniors of type B, $P(A|B)_t m_{h,t}^B$, will match with mentors of type A. Next, note that if $P(A|B)_t m_{h,t}^B$ is high enough relative to $M_{H,t}^A$, then the relative expected utility of entry for high-talent juniors of type A will be lower than $\beta\bar{Q}_t^A - o_l$ due to the crowding out effect. If this occurs, then $m_{h,t}^A = 0$ in equilibrium—high-talent juniors of type A are crowded out—which implies that $\bar{Q}_{t+1}^A = 0$. The result then follows by induction since $m_{h,t+1}^A = 0$ if $\bar{Q}_{t+1}^A = 0$.

To complete the proof, we show that if $M^{A*} < M'$ for some M', then $g(\bar{Q}_t^A, P(A|B)_t m_{h,t}^B) - o_h < \beta\bar{Q}_t^A - o_l$. First, note that $m_{h,t}^B$ is bounded from zero given $\bar{Q}_t^B > 0$, which implies that $P(A|B)_t m_{h,t}^B > \delta$ for some $\delta > 0$ for all values of M^{A*}.

Next, $\bar{Q}_t^A \to 0$ as $M^{A*} \to 0$, which implies that:

$$\lim_{M^{A*} \to 0} P(A|B) \to \frac{\beta\bar{Q}_t^B}{\beta\bar{Q}_t^B + \eta},$$

which is strictly greater than zero for all η. Lastly, since $g(0,\delta) - o_h < -o_l$, it follows that $g(\bar{Q}_t^A, \delta) - o_h < \beta\bar{Q}_t^A - o_l$ for all M^{A*} that are small enough, which completes the proof. $\square$

Proof of Proposition 6: We begin by showing that if $\bar{M}_t = \bar{M}_{t+1} = \bar{M}$ and $\eta = \infty$, then the dynamics of $\bar{Q}^A$ can be characterized by Equation 9. That is, $\bar{Q}_{t+1}^A(\bar{Q}_t^A) = \bar{Q}_{t+1}(\bar{Q}_t^A)$. Essentially, this is the same as proving that the dynamics of the profession are invariant to the size of the profession.

First, note that $P(A|B)$ is equal to zero in equilibrium if $\eta = \infty$, which means that the results of Lemma 1 apply and the following expression characterizes an interior value of m_h^{A*}:

$$m_h^{A'} = \lambda M_H^A \left[\frac{1 - f(\bar{Q}^A)}{(\beta\bar{Q}^A - o_l) - (f(\bar{Q}^A) - o_h)} \right], \tag{14}$$

which shows that Proposition 1 also applies.

Note that we wish to compare the dynamics of $\bar{Q}_t^A$ to the dynamics of $\bar{Q}_t$ in a profession without homophily at a point with $\bar{Q}_t = \bar{Q}_t^A$. At this point $M_{H,t}^A = \bar{M}M_{H,t}$, and Expression 14 gives us the following expression for $m_{h,t}^{A'}$:

$$m_{h,t}^{A'} = m'_{h,t}\bar{M},$$

where $m'_{h,t}$ is the interior value of m_h^* in the profession without homophily and $\bar{Q} = \bar{Q}^A$.

Next, note that $g(\bar{Q}_t^A, m_{h,t}^A) = g(\bar{Q}_t^A, m_{h,t})$ since:

$$\begin{aligned} g(\bar{Q}_t^A, m_{h,t}^A) &= \frac{\lambda M_{H,t}^A}{m_{h,t}^A} + \left(1 - \frac{\lambda M_{H,t}^A}{m_{h,t}^A}\right) f(\bar{Q}_t^A) \\ &= \frac{\lambda M_{H,t}\bar{M}}{m_{h,t}\bar{M}} + \left(1 - \frac{\lambda M_{H,t}\bar{M}}{m_{h,t}\bar{M}}\right) f(\bar{Q}_t^A) \\ &= g(\bar{Q}_t^A, m_{h,t}). \end{aligned}$$

Lastly, using the same steps we used to derive $\bar{Q}_{t+1}(\bar{Q}_t)$, we get:

$$\bar{Q}_{t+1}^A(\bar{Q}_t^A) = \frac{\bar{M}}{2m_{h,t}^A g(\bar{Q}_t^A, m_{h,t}^A)} = \frac{1}{2m_{h,t} g(\bar{Q}_t^A, m_{h,t})} = \bar{Q}_{t+1}(\bar{Q}_t^A).$$

That is, given a constant quota and $\eta = \infty$, the dynamics of $\bar{Q}^A$ are equivalent to the dynamics of $\bar{Q}$ given $\eta = \infty$.

In turn, this shows that given a constant quota, $\bar{M}$, $\bar{Q}^A$ is asymptotically stable at $\bar{Q}^A = \bar{Q}^*$, which allows us to characterize the following dynamic path of quotas that transition to $M^A = M^B$ and $\bar{Q}^A = \bar{Q}^*$. First, take $\bar{Q}'$ such that $\bar{Q}' < \bar{Q}^*$ and $|\bar{Q}', \bar{Q}^*| < \delta$, where $\delta > 0$ is small enough so that $\lim_{t\to\infty} \bar{Q}_t = \bar{Q}^*$. Next, take $\bar{Q}'' \in (\bar{Q}', \bar{Q}^*)$, and take n to equal the number of periods it takes the profession to transition $\bar{Q}^A$ from $\bar{Q}'$ to a point greater or equal to $\bar{Q}''$ (n is finite since $\bar{Q}^*$ is asymptotically stable).

The following algorithm results in a transition:

1. At $t = 0$, set $\bar{M}_0$ so that $\bar{Q}_1^A = \bar{Q}'$ if this implies $\bar{M}_0 < 1/2$. Otherwise set $\bar{M}_t = 1/2$ for all t.

2. Set $\bar{M}_t = \bar{M}_0$ for n periods.

3. At $t = n+1$, repeat step 1-2 and continue until $\bar{M}_0 \geq 1/2$.

Note that this algorithm will result in a transition to $M^A = M^B$ and $\bar{Q}^A = \bar{Q}^*$, but could result in a transition to $\lim_{t\to\infty} \bar{Q}_t^B > \bar{Q}^*$. However, since $\bar{Q}^*$ is stable from above and below, a transition to $M^A = M^B$ and $\bar{Q}^A = \bar{Q}^A = \bar{Q}^*$ can be achieved for a low enough δ. □

Proof of Corollary 3: Note that if the quota on mentors is gradual enough so that all juniors prefer to match with mentors of the same identity (which is effectively a restriction on δ in the proof of Proposition 6), and a quota on juniors is set so that the following holds for all t:

$$\frac{m_t^A}{m_t^B} = \frac{M_t^A}{M_t^B}.$$

Then $P(A|B)_t = 0$ for all t, and the result of Proposition 6 applies straightforwardly.

Proof of Proposition 7: Note that under $\bar{\mathbf{M}}'$, $m_{h,t}^A < m_{h,t-1}^A$ by Equation **??** given the condition that $\beta(\bar{Q}^* - \bar{Q}_{t+1}^A) < f(\bar{Q}^*) - f(\bar{Q}_{t+1}^A)$. The result then follows since a 1:1 quota will have a larger impact on perceived quality in period $t+1$ under $\bar{\mathbf{M}}'$ since it is more binding than the 1:1 quota in period t under $\bar{\mathbf{M}}$. □