

# Mentoring and Diversity

Susan Athey      Christopher Avery      Peter Zemsky\*

This revision: October, 1998

## Abstract

We study how diversity evolves at a firm with entry-level and upper-level employees who vary in ability and “type” (gender or ethnicity). The ability of entry-level employees is increased by mentoring. An employee receives more mentoring when more upper-level employees have the same type. Optimal promotions are biased by type, and this bias may favor either the minority or the majority. We characterize possible steady states, including a “glass ceiling,” where the upper level remains less diverse than the entry level. A firm may have multiple steady states, whereby temporary affirmative action policies have a long run impact.

---

\*MIT, Kennedy School of Government, and INSEAD. We would like to thank Preston McAfee and three anonymous referees for insightful comments which greatly improved the paper. We are also grateful to Jim Baron, Jeremy Bulow, Peter Diamond, Joshua Gans, Ed Lazear, Michael Kremer, Bill Lovejoy, Meg Meyer, Bob Marshall, John Roberts, Scott Stern, Bob Wilson, and seminar participants at Oxford, Stanford and Yale for helpful discussions. All errors remain our own. The authors acknowledge support from the State Farm Companies Foundation and the National Science Foundation (Athey, SBR-9631760). Athey and Avery are also grateful to the Cowles Foundation at Yale for hospitality and support while this work was being completed.

# 1 Introduction

The related topics of workplace diversity and of biased hiring and promotion decisions engender impassioned discourse in contemporary American society. Some complain that affirmative action policies lead firms to discriminate against white males. Others counter that the historic domination of management by white males puts other groups at a disadvantage, arguing that the underlying nature of the workplace inherently favors those who are from similar backgrounds as their managers. They point to the fact that in many industries and occupations, women and minorities have moved only slowly into upper-level positions, despite the fact that the labor pool and lower levels of the workforce have been diverse for some time.<sup>1</sup> This phenomenon has been referred to as the “glass ceiling.”

Some firms have made active efforts to reshape themselves, instituting affirmative action programs and hiring diversity managers. For example, IBM is known for its affirmative action program, while American Airlines, AT&T, Colgate-Palmolive Corp., DuPont Corp., and Pacific Bell have instituted diversity programs.<sup>2</sup> Other firms aggressively fight external constraints on their staffing decisions. It is difficult to reconcile this heterogeneity among firms without specifying why firms care about diversity.

This paper develops a model to analyze the relationship between the diversity of a firm’s upper level and its internal promotion policies. We consider the dynamic problem faced by a firm in choosing which of its lower level employees to promote as its existing upper-level workers retire. Employees are characterized by an initial ability for upper level work and by a type, which can be interpreted as gender, ethnicity, cultural background, personality type or even skill set (e.g. operations versus marketing skills for managers, theorists versus empiricists for academics). We assume that the lower level is split evenly between two types. We refer to the type that has most of the upper-level positions as the “majority” and the other type as the “minority.”

Employees in our model augment their initial ability by acquiring specific human capital in mentoring interactions with upper-level employees. We interpret mentoring broadly, including activities such as information sharing, informal teaching, or career advice provided by more senior workers. A critical assumption in our model is that an entry-level employee acquires more human capital from mentoring when the firm has more upper-level employees who match her type.<sup>3</sup>

---

<sup>1</sup>In many occupations, the 1980 Census revealed a higher proportion of women, Blacks, and Hispanics in “employee” positions than in “supervisor” positions (Rothstein, 1997). Large gaps exist between the proportion of women at lower levels and higher levels of large corporations (Morrison and VanGlinow, 1990) and law firms (Spurr, 1990), and also to a lesser extent in academic economics departments (Bartlett, 1997).

<sup>2</sup>*The Wall Street Journal* (August 9, 1991, p. B1).

<sup>3</sup>Our model can also be used to study the dynamically optimal *hiring* policy of a firm where the productivity of new hires depends in part on the proportion of more senior workers matching their type. Under this interpretation, the analysis addresses the evolution of the diversity of a firm’s entire

Our assumption of “type-based” mentoring is consistent with evidence from a variety of sources. Psychologists and sociologists have documented that mentoring relationships within firms are more likely to form between members of the same group.<sup>4</sup> Ibarra (1992) demonstrates that the structure of social networks depends on gender and race. More generally, communication, and thus mentoring, may be more natural and more effective when people share common interests (such as sports), cultural experiences, language, or when people have significant interactions in a community outside the workplace.<sup>5</sup> Type-biased mentoring has also been highlighted in accounts of professional partnerships. For example, a *National Law Journal* article<sup>6</sup> reported, “Despite women’s progress, the partnership ranks are 86.4% male. This puts women at a disadvantage when it comes to mentoring, the most important factor in becoming a partner. Often it’s harder for women to find a mentor, because the massive majority of partners are men, and men tend to be more comfortable mentoring other men. Rainmaking and client development—skills typically learned from a mentor—are keys to partnership.”<sup>7</sup>

We show that type-based mentoring has significant and sometimes complex effects on optimal promotion policies and on the evolution of diversity at a firm. The direct effect of type-based mentoring is that entry-level employees of the majority type acquire more human capital, and thus firms who base promotions solely on ability promote more majority employees. However, since upper-level diversity affects a firm’s profits through the mentoring of future workers, the optimal policy of a far-sighted firm will generally involve promoting workers who do not have the highest ability in order to influence the evolution of diversity over time. Thus, our approach suggests that observed differences in promotion decisions result from (i) true productivity differences, which arise (partly) as a result of a firm’s past promotion decisions, and (ii) what we call

---

workforce.

<sup>4</sup>Morrison and Van Glinow (1990) and Noe (1988) review the theory and evidence in favor of type-biased mentoring. More recently, Dreher and Cox (1996) document such differences in a survey of MBA graduates. Kanter’s classic (1977) analysis of gender roles in organizations takes a similar view, observing that “. . . numbers—proportional representation are important not only because they symbolize the presence or absence of discrimination but also because they have real consequences for performance” (p. 6).

<sup>5</sup>In a related set of historical examples, Greif (1993, 1994) shows that cooperation in trading relationships can at times be more easily sustained between members of an extended family or community.

<sup>6</sup>“Women’s Progress Slows at Top Firms,” May 6, 1996, p.A1.

<sup>7</sup>Similar claims have been made about the management consulting industry: “The lack of role models and mentors who are black is a problem...the same issues that face blacks in corporate America—lack of connections, lack of mentors, and the often-present glass ceiling—face blacks in management consulting” (Prakash, Gautam, *The Insider’s Guide to Management Consulting*, San Francisco: Wet Feet Press, 1995). Similarly, *The Wall Street Journal* (“Women Make Strides, But Men Stay Firmly in Top Company Jobs,” March 29, 1994, p. A1), reporting on the effects of culture in large U.S. companies, argued that “as long as the percentages of male managers remain high, the culture remains mostly male and, women say, indifferent or hostile to their advancement. Women say they are ignored, not taken seriously...[so] there is little chance of breaking through the glass ceiling.”

*bias* in promotions, that is, a decision to pass over some workers with higher ability as part of a long-term plan to move to a desired level of diversity. While much of the political debate about affirmative action and discrimination centers on issues of *fairness* in the evaluation of individual workers at a given point in time, our model suggests that there are important *efficiency* considerations as well in that firms can benefit from considering today's promotion decisions as an investment in future mentoring.

An important question is whether far-sighted firms benefit from increasing or decreasing diversity. In our model, the opportunity cost of a homogeneous upper level is an inability to take advantage of the scarce talent of entry-level minority employees. Since minorities receive relatively little mentoring, even a minority with a very high initial ability may be passed over in favor of a less initially able, but better mentored, majority type. Formally, we show that the scarcer is initial ability (i.e. the faster the initial ability of each type of worker diminishes in the number of that type promoted), the more the firm shifts the bias toward the minority type. Consistent with this analysis, the desire to exploit the talents of an increasingly diverse labor market is often cited by firms that actively promote workforce diversity.<sup>8</sup> Scarcity is likely to be a factor when workers use specialized skills, for example, academics, athletes, high-tech managers, or specialist physicians.

In general, we find that the optimal bias need *not* favor the minority, even if there are decreasing returns to having more mentors of a given type. Because majority employees are better mentored, their promotion rates can be higher than those of minorities, leading the firm to care more about the effective mentoring of majority than minority employees. As a result, a profit-maximizing firm will only bias its promotions to favor increased diversity if there are *sufficiently* decreasing returns to mentors of a given type.

After characterizing the optimal bias we consider the effect of type-based mentoring on the long-run level of diversity. We show that diversity of the upper level converges to a steady state, which can range from full diversity to complete homogeneity. Thus, our model can exhibit a glass ceiling phenomena where an initially homogeneous management partially diversifies, but the “invisible barrier” of type-based mentoring stalls the progress of minorities before the diversity of the upper level mirrors the lower level. We show that there may be a unique long-run level of diversity. However, we also show that there can be multiple steady state levels of diversity, whereby short-run pressure on the firm to diversify can have long-lasting impact by moving the firm from one steady state

---

<sup>8</sup>For example, *Business Week* (“White, Male, and Worried,” January 31, 1994, pp. 50-55) reported that to some companies, “managing diversity...is a competitive weapon that helps the company capitalize on its talent pool. The goal: to create a culture that enables all employees to contribute their full potential to the company's success. One way is to groom more qualified women and minorities through active succession planning.” Similarly, it has been claimed that companies use diversity programs to “attract and retain the best and the brightest...giving a headstart in recruiting and managing the workers of tomorrow.” (*USA Today*, “Setting diversity's foundation in the bottom line,” by Del Jones, October 15, 1996, p. 4B.)

to another. Type-based mentoring produces multiple steady-states because the gains of minorities are self-reinforcing: as minority representation increases in the upper level, the mentoring disadvantage faced by new minority employees decreases, making it more attractive to promote them.

Our model assumes that the lower level of the firm is heterogeneous. This assumption highlights the central tradeoff in the model: homogeneity increases the productivity of majority mentoring, but imposes an opportunity cost when the firm passes over scarce minority talent. This model is well suited to address the empirical question of why the upper levels of firms remain homogeneous even when the lower levels are much more diverse. However, it remains to consider when lower-level heterogeneity is consistent with type-biased mentoring. In Section 2, we discuss a variety of (unmodelled) labor market frictions. Further, in Section 5 we consider a static, multi-firm extension of our model that incorporates a labor market for entry-level employees. We show that when firms are initially similar, the equilibrium outcome entails at least partial diversity in the entry level. This follows because majority-type employees are a scarce resource with declining marginal value at a given firm (because the probability of promotion for a given majority-type employee falls with the number of majority employees in the entry level) and hence it can be efficient to share that resource across firms.

Our paper differs from previous theoretical work on discrimination (e.g. Arrow, 1973; Coate and Loury, 1992; Cornell and Welch, 1996; Rosen, 1997) in several ways. First, prior work is best suited for understanding discrimination in hiring rather than in promotion decisions. For example, the assumptions of incomplete or asymmetric information about worker abilities that underlies much prior work are less palatable when an employee has been with a firm for a long period of time. Second, prior work does not consider dynamically optimal policies. Finally, most studies do not unravel the forces which lead to discrimination within a firm, and all prior work deals with a one-on-one relationship between the worker and the firm. None formalizes the idea that the current diversity of a firm affects the career paths of new employees.<sup>9</sup>

An important policy implication in the received literature is that affirmative action policies can actually impede the progress of minorities by reinforcing beliefs that minorities are less qualified (Coate and Loury (1992)). Here we show that affirmative action may increase diversity in the long run by inducing the firm to shift from one steady state level of diversity to another. It is important to recognize, however, that our formal analysis involves optimizing behavior without externalities, and hence there are no inefficiencies. Unless the social welfare function includes a taste for diversity,

---

<sup>9</sup>Consider, for example, Cornell and Welch (1996) who study “screening discrimination” where employers are better able to evaluate the abilities of job applicants of the same type as themselves. Their theory depends on uncertainty about ability; their firm consists of a single employer and a single worker; and they do not consider far-sighted hiring policies where firms hire minorities today so as to better screen them in the future.

our results do not motivate government constraints on the promotion decisions of firms, such as the short-run pressure for diversity just described. There are a variety of reasons that a society might care about the diversity of firms. For example, the firms or the agents responsible for hiring might discount the future excessively, perhaps due to agency problems. Further, as we discuss in Section 6, firms may not internalize the effect of their promotion decisions on the incentives of workers to acquire human capital.<sup>10</sup>

The paper proceeds as follows. We introduce and discuss the formal model in Section 2. Section 3 characterizes when the optimal bias favors diversity or homogeneity. Section 4 analyzes how diversity evolves over time. Section 5 extends the model to include a labor market for entry-level employees, and Section 6 considers the possible effects of *ex ante* human capital acquisition. Section 7 concludes.

## 2 The Model

A single firm employs a continuum of upper-level and lower-level employees.<sup>11</sup> We normalize the measure of upper-level employees to 1. There are two types of employees, labeled *A* and *B*. The proportion of upper-level employees of type *A* in period  $t$  is denoted  $m^t$ . In each period a proportion  $r \in (0, 1]$  of the upper-level employees retire and must be replaced. Retirement rates are uniform, meaning that  $rm^t$  workers of type *A* and  $r(1 - m^t)$  workers of type *B* retire in period  $t$ . We will typically treat *A* as the majority and *B* as the minority in our exposition (i.e.,  $m^t \geq \frac{1}{2}$ ).

The firm operates an internal labor market in which retiring employees are replaced from the lower level, which has an equal number of *A*'s and *B*'s. With this assumption we introduce a cost of upper-level homogeneity which we believe to be quite general: homogeneous firms bear the opportunity cost of passing over talented workers of the opposite type within their organization. There are numerous reasons why the lower levels of organizations are diverse despite type-based mentoring (which introduces a benefit to segregating entry levels by type). For example, firms generally differ in their locations, cultures, and skill requirements, and thus workers may have firm-specific preferences and skills which make it beneficial for a diverse group of employees to work at the same firm. We formalize such costs in Section 5. Search costs might reinforce the effects of firm-worker matching. Further, in the present legal and political environment,

---

<sup>10</sup>Yet another potential source of inefficiencies is that successful members of a type (i.e. individuals who attain upper level positions) may serve as “role models” for younger members of their community. This role modeling would be very much like our broad definition of mentoring except that the human capital transfers would happen *outside* of the firm and hence would not be factored into firm decision making.

<sup>11</sup>In our 1994 Stanford GSB working paper we consider a model of type-based mentoring with a finite number of employees where initial ability is stochastic. The qualitative insights of the stochastic model are the same, but the deterministic model analyzed in this paper is more tractable.

a variety of non-market forces create pressure for diverse lower levels.<sup>12</sup> Finally, Section 5 shows that partial diversity of the lower level can arise in equilibrium even without labor market frictions.

We assume that there are at least  $r$  entry-level workers of each type and that the two pools of lower-level workers are symmetric in their initial abilities. Member  $\theta \in [0, r]$  of a pool of entry-level workers has an initial ability  $x(\theta)$ , which represents the surplus received by the firm when worker  $\theta$  takes an upper-level position. There are several ways to justify the presence of such a surplus. The worker’s ability might be specific to the firm-worker match, it might not be observable by other firms, or there might be frictions in the labor market.

The function  $x(\theta)$  is assumed to be nonincreasing. Thus, the quality of the marginal worker from each pool (weakly) decreases as the firm digs deeper into that pool, representing scarcity of ability for upper-level work. Such scarcity may be especially relevant for jobs in the service sector, jobs requiring education and specialized skills, jobs where “stars” are important (such as academics, entertainment, and sports), and high-level management positions in firms.

A worker’s type and initial ability are observed by the employer. Each entry-level employee gains additional (firm-specific) skills required for upper-level work through mentoring, represented by the function  $\mu(\cdot)$ , which depends on the proportion of upper-level employees with the same type.<sup>13</sup> This mentoring function is assumed to be increasing and continuously differentiable. The overall (lifetime) contribution to the firm’s profit from promoting an applicant, her “surplus”, is the following function of her type, her index and the composition of the firm when she is promoted:<sup>14</sup>

$$\begin{aligned} s_A(\theta, m) &= x(\theta) + \mu(m) \text{ and} \\ s_B(\theta, m) &= x(\theta) + \mu(1 - m). \end{aligned}$$

We denote by  $z^t \in [0, r]$  the measure of the promoted employees in period  $t$  who are of type  $A$ ;  $r - z^t$  are then of type  $B$ . The firm’s per-period profit function is the total

---

<sup>12</sup>Segregated firms may be at risk of government and civil litigation for biased hiring practices. They may also face social pressure to mirror the composition of their local labor markets.

<sup>13</sup>The effectiveness of mentoring does not depend on the ability of upper-level employees. If mentoring depends linearly on the average ability of upper-level employees, then our analysis is unaffected. However, the analysis *is* affected if mentoring depends on the average ability of upper-level employees with the same type. We expect that such dependence would reduce the magnitude of the promotion bias because promoting less able candidates becomes less attractive.

<sup>14</sup>The surplus functions do not include the indirect contribution from future mentoring of other workers; these effects are incorporated in the firm’s dynamic programming problem that follows. Further, observe that while our primary interpretation of mentoring is firm-specific, any general skills in mentoring may accrue to the workers. Thus the surplus functions can be interpreted as the benefit to the firm net of wages.

surplus generated by its new upper-level employees:

$$\pi(m, z) = \int_0^z s_A(\theta, m) d\theta + \int_0^{r-z} s_B(\theta, m) d\theta.$$

The firm seeks to maximize the discounted sum of its per-period profits by choosing a sequence of promotion policies  $(z^1, z^2, \dots)$  to solve the following maximization problem:

$$\begin{aligned} V(m) = & \max_{(z^1, z^2, \dots)} \sum_{t=0}^{\infty} \delta^t \pi(m^t, z^t) \\ & \text{s.t. } m^{t+1} = (1-r)m^t + z^t, \\ & z^t \in [0, r] \text{ and } m^0 = m. \end{aligned} \tag{1}$$

Given a value function (which is established to be well-behaved for our model in the Appendix), we can define the optimal policy correspondence by

$$z^*(m) = \{z \in [0, r] | V(m) = \pi(m, z) + \delta V(z + (1-r)m)\}.$$

We are interested in how  $m^t$  evolves over time and in how the firm biases its promotions. We define the *unbiased* (or myopic) promotion policy  $z^{UB}(m)$  implicitly from the equation

$$s_A(z^{UB}, m) = s_B(r - z^{UB}, m),$$

which equalizes the contribution of the marginal promoted employee of each type. If the worst type  $A$  is better than the best type  $B$  ( $s_A(r, m) > s_B(0, m)$ ) so that the marginal contributions cannot be equalized, then the firm only promotes type  $A$ 's and  $z^{UB}(m) = r$ . Similarly, if  $s_B(r, m) > s_A(0, m)$  then  $z^{UB}(m) = 0$ .

We can now define the bias in the firm's optimal promotion policy as

$$b(m) = z^{UB}(m) - z^*(m).$$

A positive bias  $b(m) > 0$  is then a bias in favor of type  $B$ .<sup>15</sup>

---

<sup>15</sup>There are other ways to define bias in our model. Given the identical initial ability of each type, one could define bias as any deviation from equal promotion rates (i.e. as  $|z^*(m) - r/2|$ ). This definition produces more of a bias for the majority. Alternatively, one could define the bias as any deviation from promoting the applicant who contributes the most to long-run firm profits, including the indirect contribution from mentoring others. Under this definition, there is no bias. One advantage of our definition is that it is linked to a firm's preferences about diversity. Thus a firm which biases for the minority (by our definition) would also be willing to incur other costs to facilitate minority career advancement such as hiring diversity managers.



Given the promotion policies  $z^{UB}$  and  $z^*$ , we can define transition functions  $M^{UB}(m)$  and  $M^*(m)$  which give the state achieved in period  $t + 1$  as a function of the period  $t$  state,  $m$ :

$$M^{UB}(m) = m(1 - r) + z^{UB}(m).$$

$$M^*(m) = \{m' | m' = (1 - r)m + z, \text{ where } z \in z^*(m)\}.$$

Figure 1 plots the surplus functions,  $s_A(\theta, m)$  and  $s_B(\theta, m)$ , for an example. Since type  $A$  is the majority ( $m = .9$ ), the surplus function is higher by the amount of the mentoring differential, in this case  $\mu(.9) - \mu(.1)$ . When unbiased promotion policies are used, by definition, the surplus of the marginal promoted employee is the same across the worker types, so that the firm promotes mostly type  $A$  workers. In this example, the optimal bias is positive, so that when the optimal promotion policy is used, fewer type  $A$  workers are promoted, and the marginal type  $A$  worker is better than the marginal type  $B$  worker.

It is possible to reinterpret our model as applying to hiring rather than promotion decisions. Then what was the “upper level” becomes the firm’s whole workforce, and what was the “lower level” becomes the pool of applicants. The firm anticipates the effects of mentoring when it makes its hiring decisions. The firm faces an exogenously diverse labor pool, which might be justified by labor market frictions or heterogeneity in the quality of the firm-worker match (independent of type).

### 3 The Optimal Promotion Policy

We begin our analysis by characterizing the direction of the optimal promotion bias. As a building block, we first characterize the effect of diversity on the profit of a firm which lives for only one period. The one-period analysis highlights the critical trade-off between maximizing the mentoring gain for one of the applicant pools and exploiting the scarce initial talent in both pools.

#### 3.1 The One Period Problem

Consider a firm which lives for only one period (i.e.  $\delta = 0$ ). The optimal policy for such a firm is unbiased promotions ( $z^{UB}(m)$ ) since the firm does not care about mentoring performed by promoted workers. The profit of such a firm is then

$$\pi^{UB}(m) = \pi(m, z^{UB}(m)).$$

Suppressing the dependence of  $z^{UB}$  on  $m$ , we can rewrite this as follows:

$$\pi^{UB}(m) = z^{UB}\mu(m) + (r - z^{UB})\mu(1 - m) + \int_0^{z^{UB}} x(\theta)d\theta + \int_0^{r - z^{UB}} x(\theta)d\theta.$$

By the envelope theorem, the effect of a change in the initial level of diversity is given by:

$$\frac{d}{dm}\pi^{UB}(m) = z^{UB}\mu'(m) - (r - z^{UB})\mu'(1 - m). \quad (2)$$

Increasing the initial proportion of type  $A$  workers leads to an increase of  $\mu'(m)$  in the mentoring received by the  $z^{UB}$  type  $A$  workers who will be promoted, as well as a corresponding reduction of  $\mu'(1 - m)$  for the  $r - z^{UB}$  type  $B$  workers who are promoted. Thus, the sign of  $d\pi^{UB}(m)/dm$  depends on the curvature of the  $\mu$  function as well as the relative proportions of type  $A$  and  $B$  workers promoted. If the mentoring function is concave, then  $\mu'(1 - m) > \mu'(m)$ , and increases in diversity (i.e. a lower  $m$ ) increase the mentoring of minority employees more than they decrease the mentoring of majority hires. However, since the mentoring function is nondecreasing, majority types gain more from mentoring than minority types, and a myopic firm promotes more of the (majority)  $A$  types than the (minority)  $B$  types ( $z^{UB} > r/2 > r - z^{UB}$ ); thus, profits are more sensitive to the mentoring of  $A$  types than  $B$  types.

As a result, the firm's profit increases with the diversity of its upper level only if the mentoring function is concave *and* the degree of concavity is sufficient to overcome the larger weight placed on mentoring outcomes for majority workers. We shall say that *sufficient concavity* (SCV) holds at  $m > \frac{1}{2}$  if  $\frac{d}{dm}\pi^{UB}(m) < 0$ , that is, if

$$\mu'(m) < \frac{r - z^{UB}(m)}{z^{UB}(m)}\mu'(1 - m). \quad (\text{SCV})$$

Further, we say that *sufficient concavity holds everywhere* (SCV everywhere) if SCV holds for all  $m > \frac{1}{2}$ . Similarly, we shall say that *sufficient concavity holds nowhere* (SCV nowhere) if  $\mu$  fails to be SCV for each  $m \geq \frac{1}{2}$ .

### 3.2 When Does Sufficient Concavity Hold?

Although derived for the one period problem, the SCV condition is useful in characterizing the promotion bias and long-run dynamics for infinitely-lived firms. Thus, we pause to discuss the curvature of the mentoring function and to characterize the role of several exogenous parameters.

In many settings, we expect that the mentoring function should exhibit decreasing returns. In its most literal sense, mentoring refers to a voluntary relationship between a more experienced employee and a new hire. An increase in the proportion of upper-level employees of a given type may allow lower level employees who would otherwise be unmatched to find a mentor (or a more appropriate mentor). While the first minority-type upper-level workers yield high returns by providing mentoring to those who would not otherwise have received it, at higher levels of diversity, additional upper-level minorities simply provide somewhat better mentoring possibilities to those

who would have been mentored anyway. However, our model does not *require* that there are decreasing returns. We can accommodate mentoring functions with constant or increasing returns, which just implies SCV nowhere. Section 4.3 contains an example with a “critical mass” mentoring function which is first convex and then concave.

In addition to the shape of the mentoring function, the retirement rate, the importance of mentoring and the scarcity of initial talent all affect whether SCV holds through their impact on  $z^{UB}(m)$ . The retirement rate  $r$  represents the rate of upper-level turnover relative to the size of the firm. Industries with low turnover at the senior level relative to the size of the firm might include law, academics, businesses organized by partnership, and large bureaucracies; high turnover occurs in many high-technology industries, as well as intermediate levels of hierarchies.

We now introduce a parameter  $\alpha$  to describe the importance of mentoring. For any mentoring function  $\hat{\mu}$ , we can define  $\mu(m) = \alpha\hat{\mu}(m)$ , and explore the effects of varying  $\alpha$ . Industries where mentoring is likely to be important relative to initial ability include services, law, academics, industries where networking and information sharing play an important role, and industries where apprenticeships are important components of training. In contrast, mentoring will be less important in industries where jobs are well-defined, require few specialized skills, individual production, and involve little information sharing.

Finally, we introduce a parameter  $\gamma$  to describe the scarcity of initial ability. An increase in  $\gamma$  makes  $x$  steeper, so that for  $z > \frac{r}{2}$ ,  $x(z; \gamma) - x(r - z; \gamma)$  is nonincreasing in  $\gamma$ . Thus, the higher is  $\gamma$ , the greater is the cost (in terms of lower average initial ability of promoted employees) of deviating from promoting  $r/2$  of each type. The linear initial ability function  $x(\theta; \gamma) = \kappa - \gamma\theta$  satisfies our assumptions. Industries where scarcity is important include industries which require specialized skills and experience, such as high-level management, or industries where “stars” are important, such as academics.

The following proposition summarizes the effects of these three parameters on  $\frac{d}{dm}\pi^{UB}(m; \alpha, \gamma, r)$ , the sign of which determines whether SCV is satisfied.

**Proposition 1** *For  $m > \frac{1}{2}$ ,  $\frac{d}{dm}\pi^{UB}(m; \alpha, \gamma, r)$  is:*

*(i) nonincreasing in  $\gamma$ ; (ii) nondecreasing in  $\alpha$ ; and (iii) nonincreasing (nondecreasing) in  $r$  whenever  $\frac{d}{dr}z^{UB}(m; r, \alpha, \gamma) \cdot \frac{r}{z} \leq (\geq) 1$ , and hence for  $x(\theta) = \kappa - \gamma\theta$  it is decreasing in  $r$ .*

First, an increase in scarcity (as parametrized by  $\gamma$ ) increases the proportion of minorities promoted, thereby increasing the benefits of upper level diversity so that SCV is more likely to hold. Conversely, an increase in the importance of mentoring (as parametrized by  $\alpha$ ) increases  $z^{UB}$ , since the majority type gains a larger advantage, and thus makes SCV less likely to hold. When  $\alpha$  becomes sufficiently large, the firm’s optimal policy will only select majority types (SCV will hold nowhere); for  $\alpha$  sufficiently close to zero, unbiased promotion selects almost the same number of majority

and minority types, and then concavity and SCV are (approximately) equivalent. Finally, an increase in  $r$  requires the firm to promote more workers, which changes the proportion of type  $A$  workers promoted. If the elasticity of  $z^{UB}$  with respect to  $r$  is less than 1, the proportion of type  $A$  workers falls when  $r$  increases, and so condition SCV is more likely to hold.

In summary, the condition SCV, which is associated with a preference for diversity, is more likely to hold the greater is the concavity of the mentoring function, the scarcer is initial ability, the less important is mentoring and the greater is turnover (assuming ability falls off at a constant rate).

### 3.3 The Infinite Horizon Problem

Building on the analysis of Section 3.1, we use an inductive argument to sign the optimal promotion bias of a far-sighted firm. Start by considering the optimal bias of a firm which will live for one more period. In the final period, the firm does not care about future mentoring, and unbiased promotions are optimal; thus, equation (2) characterizes the firm's preferences over the level of last-period diversity. Consequently, if SCV holds everywhere, the firm will bias its first period promotions in favor of the minority. Conversely, if SCV holds nowhere, the firm will bias its first period promotion in favor of the majority type.

But what if the firm lives for more than two periods? Our next step is to show that the firm's preferences about diversity in a two-period problem carry over to the infinite horizon problem. The critical feature of the model for this result is that, in terms of today's profits, the more type  $A$ 's the firm intends to promote, the higher the returns to having more type  $A$ 's in the upper level:

$$\frac{\partial^2}{\partial m \partial z} \pi(m, z) = \mu'(m) + \mu'(1 - m) > 0.$$

Consider the case where SCV holds everywhere, and suppose that the future value of the firm starting in period  $t + 1$  is nonincreasing in  $m^{t+1}$ . Looking ahead to period  $t + 1$ , where diversity is valuable, the firm in period  $t$  promotes more minorities than it would with unbiased promotions ( $z^t(m) \leq z^{UB}(m)$ ). But since  $\pi(m^t, z)$  is decreasing in  $m^t$  at  $z = z^{UB}(m^t)$  (by the definition of SCV), the fact that  $\frac{\partial^2}{\partial m \partial z} \pi \geq 0$  implies that the returns to upper-level type  $A$ 's are even lower using optimal promotions. Thus, the value of the firm in period  $t$  is nonincreasing in  $m$ . Formally:

**Proposition 2** *(i) Suppose that SCV holds everywhere. Then the value of the firm is increasing with the level of diversity ( $dV/dm < 0$ ) and the optimal promotion policy is biased in favor of the minority. (ii) Now suppose that SCV holds nowhere. Then the value of the firm is increasing as the upper level becomes more homogeneous ( $dV/dm > 0$ ) and the optimal promotion policy is biased in favor of the majority.*

Propositions 1 and 2 together establish the relationship between the parameters of the model and the sign of the optimal bias.

There are, of course, parameter values for which SCV holds in some regions but not others. Then it is possible that the firm biases in favor of the majority for some values of  $m$  and in favor of the minority for others. Consider the following example:  $\mu(m) = m^{.11}$ ,  $x(\theta) = 1 - \theta$ , and  $r = .3$ . Figure 2 plots the one-period value function,  $\pi^{UB}(m)$  for  $m \geq \frac{1}{2}$ . Note that profits increase with diversity except when  $m \geq .927$ . Figure 3 illustrates the unbiased promotion rate ( $z^{UB}$ ) for type  $A$  applicants as well as the optimal promotion rates for  $\delta = .3$  and  $\delta = .95$ . The bias in favor of type  $B$  workers,  $b(m; \delta)$ , is the difference between the unbiased promotion rate ( $\delta = 0$ ) for type  $A$ 's and the biased promotion rate.

Figure 3 shows that for  $\delta = .95$ , the firm always biases in favor of the minority; despite the fact that the one-period value function is increasing near 1, the firm finds it worthwhile to bias promotions towards full diversity, which maximizes the one-period value function. In contrast, if  $\delta = .3$ , the firm's optimal policy is to bias in favor of the majority in the region  $[.93, .96]$ , and in favor of the minority elsewhere. Even within a single firm, attitudes towards diversity will change with the initial conditions. We will further discuss the dynamics of this example in the next section.

## 4 The Dynamics of Diversity

We now analyze the evolution of diversity at our firm. Our model identifies a dynamic path of diversity levels (and associated biases) followed by a firm from any initial condition to a steady state. While we illustrate these dynamics with numerical examples, our formal results focus on the steady state levels of diversity. In Section 4.1 we prove convergence to a steady states. In Sections 4.2 and 4.3 we look at the set of steady states with and without the optimal bias, respectively. In Section 4.4 we conduct a comparative statics analysis. We shall say that the firm achieves full diversity when convergence is to  $m = \frac{1}{2}$  and we shall refer to  $|m - \frac{1}{2}|$  as the level of diversity.

### 4.1 Existence of Stable Steady States

A steady state level of diversity is defined as a fixed point in the optimal transition correspondence  $m_s \in M^*(m_s)$ , so that it is optimal for a firm with diversity  $m_s$  to maintain that level. That is, at a steady state, the measure of retiring  $A$  managers exactly equals the measure of type  $A$ 's that are promoted ( $rm_s \in z^*(m_s)$ ). Similarly, a steady state with unbiased promotions satisfies  $m_s = M^{UB}(m_s)$ . Figure 3 graphs the (uniform) retirement function and the promotion policy for three values of  $\delta$ ; each crossing of the retirement and promotion functions represents a steady state. Thus,  $\{.5, .936, 1\}$  are steady states greater than or equal to  $.5$  when  $\delta = .3$ . We shall refer to

a steady state which is always reached from a neighborhood surrounding it as a *stable steady state* (when  $\delta = .3$ , the points  $\{.5, 1\}$  in Figure 3). A *critical mass point* has no basin of attraction, and so the long-run outcome of the firm is determined by whether the firm's initial level of diversity is greater than the critical mass (when  $\delta = .3$ , the point  $\{.936\}$  in Figure 3).<sup>16</sup>

Let  $S^*$  be the set of stable steady states under the optimal promotion policy and let  $S^{UB}$  be the set of stable steady states with myopic promotions.<sup>17</sup> The following proposition proves that the level of diversity always converges to a steady state.

**Lemma 3** (i)  $M^{UB}(\cdot)$  is a single-valued, nondecreasing continuous function, and  $M^{UB}(m) \geq \frac{1}{2}$  for  $m > \frac{1}{2}$ .  
(ii)  $M^*(\cdot)$  is upper semi-continuous, and except for upward jumps, it is a single-valued, nondecreasing and continuous function. Further,  $M^*(m) \geq \frac{1}{2}$  for  $m > \frac{1}{2}$ .  
(iii) With either biased or unbiased promotions, the diversity of the firm converges to a steady state in  $[\frac{1}{2}, 1]$  for  $m^0 > \frac{1}{2}$ .

Lemma 3 establishes that, under either unbiased or optimal promotion policies, the more type A's in the upper level today, the more there will be tomorrow. Thus, the majority type never changes. These results (similar to Proposition 2) follow from the fact that  $\frac{\partial^2}{\partial m^t \partial m^{t+1}} \pi(m^t, m^{t+1} - rm^t) \geq 0$ : in terms of today's profits, the marginal return to increasing the future proportion of type A workers is nondecreasing in the proportion of type A workers in the upper level today. A transition function that is nondecreasing and continuous (but for upward jumps) is sufficient to establish convergence to a steady state.

This is the first of several results that explores the extent to which "history matters" in our model. Here we see that the majority type never becomes the minority. We now consider whether initial asymmetries are perpetuated by a glass ceiling (i.e. stable steady states other than  $m = \frac{1}{2}$ ), and whether there are multiple steady states.

## 4.2 Stable Steady States with Unbiased Promotions

We start by characterizing  $S^{UB}$ , the set of stable steady states when the firm uses an unbiased, or myopic, promotion policy. This is of interest because a variety of factors (such as legal restrictions or agency problems) may lead some organizations to pursue myopic policies, and further because it helps us to isolate the effect of the optimal bias on the set of steady states.

---

<sup>16</sup>More formally, the level of diversity  $m$  is stable (a critical mass point) for the transition function  $M$  if there exists a  $v > 0$  such that, for all  $v > \varepsilon > 0$  such that  $M(m + \varepsilon)$  and  $M(m - \varepsilon)$  are single-valued,  $m + \varepsilon \geq (<)M(m + \varepsilon)$  and  $m - \varepsilon \leq (>)M(m - \varepsilon)$ .

<sup>17</sup>We use the following notion of symmetry when analyzing stable steady states:  $S$  is symmetric if  $m \in S \Rightarrow 1 - m \in S$ . Since we have assumed that the two types are symmetric in all other ways, our firm's policies and steady states will be symmetric as well.

**Proposition 4** (i)  $\frac{1}{2} \in S^{UB}$  iff  $\mu'(\frac{1}{2}) < -rx'(\frac{r}{2})$ . (ii) For any finite, symmetric set of points  $X \subset [0, 1]$ , there exists a nonincreasing initial ability function  $x(\cdot)$  and a nondecreasing mentoring function  $\mu$  such that  $S^{UB} = X$ .

Part (i) says that full diversity ( $m = \frac{1}{2}$ ) is stable when the asymmetries created by type-based mentoring in the neighborhood of full diversity (i.e.  $\mu'(\frac{1}{2})$ ) are small relative to scarcity of talent (i.e.  $rx'(\frac{r}{2})$ ).<sup>18</sup> Part (ii) establishes the possibility of multiple stable steady states, showing that  $S^{UB}$  can take (almost) any form long as it is symmetric. This extreme result uses very particular functional forms. For many functional forms we expect that there will be no more than three steady states. Consider the following example:

**Example 1** Suppose the initial ability function is linear,  $x(\theta) = \kappa - \gamma\theta$ . If  $\mu'''(m) > 0 \forall m$ , then  $S^{UB} \subset \{0, \frac{1}{2}, 1\}$ . If  $\mu'''(m) < 0 \forall m$ , then  $S^{UB} = \{m_s, 1 - m_s\}$  for some  $m_s \geq \frac{1}{2}$ .

When the initial ability function is linear and  $\mu'''(m) > 0$ , the transition function  $M^{UB}$  is convex; thus, firms who simply hire the most “able” workers will eventually attain either full diversity or full homogeneity. In contrast, if  $\mu'''(m) < 0$ , the transition function  $M^{UB}$  is concave, and an intermediate outcome is possible. Full diversity might never be attained; but if it is attained, it is the unique long-run stable steady state.

### 4.3 Stable Steady States with Optimal Promotion Policies

We now explore the effect of the optimal bias on the evolution of diversity. We can write  $M^*(m) = M^{UB}(m) - b(m)$ , illustrating that the effect of the optimal bias is to shift the unbiased transition function. Restrict attention to the interval  $[\frac{1}{2}, 1]$ . If the bias has an unambiguous sign (as when SCV holds everywhere or nowhere), then the transition function shifts down when the bias is positive and up when it is negative. Such monotonic shifts in the transition function produce monotonic shifts in the set of stable steady states (Milgrom and Roberts, 1994).

**Definition 1** Consider two sets of stable steady states,  $S_1$  and  $S_2$ . We say that  $S_1$  is **more diverse** than  $S_2$  if

$$\begin{aligned} \min\{m | m \in S_2, m \geq \frac{1}{2}\} &\geq \min\{m | m \in S_1, m \geq \frac{1}{2}\}, \\ \max\{m | m \in S_2, m \geq \frac{1}{2}\} &\geq \max\{m | m \in S_1, m \geq \frac{1}{2}\}. \end{aligned}$$

---

<sup>18</sup>Here, as elsewhere, the scarcity of initial ability depends on the product of the retirement rate and the derivative of the initial ability function  $x$ . The retirement rate enters because it determines the rate at which the firms dips into its applicant pools.

**Proposition 5** (i) *If SCV holds everywhere, then  $S^*$  is more diverse than  $S^{UB}$ , and  $\frac{1}{2} \in S^*$  whenever  $\mu'(\frac{1}{2}) < -rx'(\frac{r}{2})$ .*  
(ii) *If SCV holds nowhere, then  $S^*$  is less diverse than  $S^{UB}$ , and  $\frac{1}{2}$  is never a stable steady state.*

Thus a bias which favors the minority (as when SCV holds everywhere) increases long-run diversity, while a bias in favor of the majority (as when SCV holds nowhere) decreases long-run diversity. As a result, the effect of public policies which seek to force firms to promote “the best person for the job” can have ambiguous effects. Full diversity is never achieved when the bias favors the majority. When the bias favors the minority, full diversity is at least one possible outcome, as long as the asymmetries created by type-based mentoring are not too large relative to the scarcity of talent.

**Example 2** *Suppose the initial ability function is linear,  $x(\theta) = \kappa - \gamma\theta$ . If  $\mu'''(m) < 0 \forall m$  and  $\mu'(\frac{1}{2}) > r\gamma$ , then there exists  $\delta$  small enough that  $S^* = \{m_s, 1 - m_s\}$  for some  $m_s > \frac{1}{2}$ .*

Example 2 and Proposition 5 (ii) identify two situations in which our model exhibits the glass ceiling phenomena. When full diversity is not a stable steady state, firms which start with homogeneous management may diversify initially, but the progress of minorities always stalls: the “invisible barrier” of type-based mentoring keeps minority representation in the upper level of the organization from mirroring the diversity at the lower level.

In Example 2, historically given initial conditions determine which type is in the majority, but the long-run level of diversity is unique. Figure 3 with  $\delta = .3$  gives an example with multiple steady states greater or equal to  $\frac{1}{2}$ .<sup>19</sup> In such cases, initial conditions also determine the level of long-run diversity. With  $\delta = .3$ , the firm converges to homogeneity if  $m^0 > .936$ , while otherwise convergence is to full diversity. One implication of these dynamics is that short-lived pressure on a firm to diversify (e.g. from legislation or other forms of social pressure) can have long lasting impact on the level of diversity. When pressure pushes diversity above a critical mass point, the gains of the minorities are self-reinforcing even if the pressure is removed.

Recall that Proposition 2 shows how firm profit varies with the level of diversity. Thus, we have shown that firms employing the same technologies and facing identical worker pools can have different levels of profitability depending on their historical conditions. While the less profitable firm could imitate the organization of the other firm, the cost of promoting less able employees in order to influence to evolution of diversity is greater than the benefit. Because the optimal bias depends on the level of

---

<sup>19</sup>Further, it can be shown that even if SCV holds everywhere (which is not the case in Figure 3), there may exist multiple stable steady states  $m_s^H > m_s^L \geq 1/2$ . For example, if  $\mu(m) = .1m^1$ ,  $x(\theta) = 1 - \theta$ ,  $r = .05$ , an example can be constructed with the highest stable state being  $m = 1$ .



diversity, this implies that firms with the same technology and worker pools can vary in their attitudes towards diversity.

In addition to variation across firms, our theory also predicts variation in diversity and bias at a firm over time. Figure 3 with  $\delta = .3$  illustrates such variation. A firm starting just to the left of the critical mass point  $m^0 = .936$  initially biases in favor of the majority, but as diversity increases the bias shifts to favor the minority.<sup>20</sup> A firm starting to the right of  $m^0 = .936$  initially biases in favor of the majority, but as homogeneity increases the optimal bias shifts to favor the minority.

A second example is illustrated in Figures 4 and 5. Figure 4 illustrates a critical mass mentoring function, where mentoring is relatively ineffective until the minorities reach 20% of the firm.<sup>21</sup> Such a mentoring function is associated with a glass ceiling and multiple steady states. For the mentoring function of Figure 4 with  $r = .3$  and  $x(\theta) = 1 - \theta$ , unbiased or short-sighted promotion policies ( $\delta = 0$ ) lead to stable steady states at  $S^{UB} = \{.1, .5, .9\}$  even though profits are maximized at full diversity. A sufficiently far-sighted firm ( $\delta = .95$ ) biases promotions so that  $S^* = \{\frac{1}{2}\}$ . Figure 5 graphs the proportion of type *A* managers in each period for  $\delta = .3$ , given a variety of different initial conditions. There is a glass ceiling, because starting from homogeneity, minorities make some initial progress into the firm; however, they never reach a level of representation above 10%. The increasing returns in a critical mass mentoring strengthen the general tendency for minority gains to be self-perpetuating after the critical mass is reached.

## 4.4 Comparative Statics

We now consider comparative statics on the set of steady states with respect to the scarcity of initial ability, the discount factor, the importance of mentoring and the retirement rate. We use the parameters introduced in Section 3.2.

We take a local approach to comparative statics in that we look at comparative statics for states  $m_s$  which are steady states.<sup>22</sup> Despite limits of this approach, we still find that several of our results from Proposition 1 must be qualified in the fully

---

<sup>20</sup>Diversity increases because the optimal bias and type-based mentoring are not sufficiently strong to overcome the greater retirement rate of the majority type.

<sup>21</sup>A critical mass mentoring function is one which is first convex and then concave. This mentoring function incorporates the idea that  $\mu(m)$  increases slowly at first due to the relative ineffectiveness of mentors who are themselves highly isolated in the organization. Mentoring only becomes effective once a critical mass of mentors is reached. Such curvature could also arise if almost complete homogeneity in a group greatly increases the efficiency of communication and information sharing so that the last few steps toward homogeneity greatly enhance mentoring.

<sup>22</sup>Another approach would be to derive comparative statics on the optimal bias, and use these results to analyze changes in steady states. However, there is no necessary connection between comparative statics on the bias from a two-period model (which would follow from Proposition 1), and comparative statics on the bias in the infinite horizon model.

dynamic problem. We begin by examining the First Order Conditions (FOC) for the unique optimal promotion rule which must be satisfied at a stable steady state:

**Lemma 6** *For all  $m_s \neq \frac{1}{2}$  which are stable steady states, there is a unique optimal promotion policy,  $z^*(m_s)$ . An interior steady state  $m_s \notin \{0, \frac{1}{2}, 1\}$  satisfies the following first order condition*

$$FOC_s(m_s) = \pi_2(m_s, rm_s) + \frac{\delta}{1 - (1 - r)\delta} \pi_1(m_s, rm_s) = 0.$$

We see that the return to promoting more majority type workers today is balanced against the (weighted) return to having a lower level of diversity in the next period. The following Proposition summarizes the main comparative statics results.

**Proposition 7** *Each stable steady state  $m_s \geq \frac{1}{2}$  is:*

- (i) *Nonincreasing in response to a small increase in  $\gamma$ .*
- (ii) *When SCV holds everywhere (nowhere), nonincreasing (nondecreasing) in response to a small increase in  $\delta$ .*
- (iii) (a) *When SCV holds nowhere, nondecreasing in response to a small increase in  $\alpha$ .* (b) *When  $\delta$  is sufficiently small, nondecreasing in  $\alpha$ .*
- (iv) (a) *When SCV holds everywhere and  $x(\theta) = \kappa - \gamma\theta$ , nonincreasing in response to a small increase in  $r$ .* (b) *When  $\delta$  is sufficiently small, nonincreasing (nondecreasing) in  $r$  if  $\partial z^{UB} / \partial r < (>) m_s$ .*

In interpreting the results, it is useful to recall that  $M^*(m) = M^{UB}(m) - b(m)$ . That is, the effect of a parameter can be decomposed into its effect on the unbiased transition and the effect on the optimal bias, with the effect on the unbiased transition already characterized in Proposition 1.

Part (i) of Proposition 7 shows that the faster initial ability declines with the proportion of a given type promoted, the more diverse the upper level will be in steady state. Note that the direct effect of a shift in  $\gamma$  on the unbiased dynamics—more promoted minorities—is reinforced by the effect of  $\gamma$  on the optimal bias. As the firm is promoting more minorities, it cares more about their mentoring and hence the change in the bias favors diversity.

The parameter  $\delta$  has a simple effect in our model. As the future looms larger, the optimal bias becomes larger, since the firm cares more about the effectiveness of future mentoring. Hence, the effect on the dynamics depends on the direction of the optimal bias. If, for example, the bias is for the minority, then diversity increases. This result identifies the nature of the inefficiency which arises if there is excessive discounting by the agents responsible for promotion decisions. In practice, long-term compensation schemes are rarely used for agents in the middle levels of a firm's hierarchy, and workers

take into account the fact that they will leave the firm (or at least their current job) with positive probability in the future. Then a firm with a SCV everywhere mentoring function settles for too little diversity. Inefficiencies of this kind are consistent with the observation that many firms have internal rules or policies which restrict decisions about diversity in hiring and promotion.<sup>23</sup> Furthermore, since such agency problems may exist even in top management, shareholders may wish to constrain the decisions of top management about diversity.<sup>24</sup>

The comparative statics for  $\alpha$  and  $r$  require additional assumptions. Consider first the effect of  $\alpha$ , where  $\mu(m) = \alpha\hat{\mu}(m)$ . The direct effect on the unbiased dynamics is a shift towards greater homogeneity as the mentoring disadvantage of minorities increases. However, unlike with  $\gamma$ , the shift in the unbiased dynamics need not carry over to the optimal bias. As with  $\delta$ , an increase in  $\alpha$  makes the firm more willing to bias promotions to optimize mentoring, which is now more important. If the mentoring function is everywhere SCV, then the three effects go in the same direction and we get the expected result that diversity falls with  $\alpha$ . Also, for  $\delta$  sufficiently small, the bias is small enough in magnitude that the direct effect dominates.

Now consider the rate of turnover. As the retirement rate goes up, the firm must dig deeper into its applicant pool, which has an ambiguous effect on the unbiased dynamics. Thus for  $\delta$  small, the direction of the effect depends on the elasticity of the unbiased promotion rate. However, these results do not necessarily generalize. As with  $\delta$ , increases in  $r$  make the firm more willing to bias promotions (in either direction) since current promotions have an immediate and large effect on diversity. Therefore, when the initial ability function is linear (and thus faster turnover leads to a higher proportion of minority promotions by a myopic firm), unambiguous predictions about steady states follow only when the bias favors the minority.

Finally, we consider briefly asymmetries in the initial ability function, which may be of particular interest for policy applications given observed differences in the prevalence and education levels of different types in the population. Although we do not formally model asymmetries, the effect is straightforward given our results about scarcity ( $\gamma$ ). As the initial ability of one type gets less scarce, the promotion rates and bias shift

---

<sup>23</sup>Our theory also raises the possibility that firms with agency problems want to adopt rules restricting diversity, which we do not typically observe, probably because of legal and social costs associated with explicit discrimination against minorities.

<sup>24</sup>The ability of shareholders to dictate diversity policy has been a topic of recent debate by the Securities and Exchange Commission (SEC) (see Wallman, Steven, "Equality Is More Than 'Ordinary Business,'" *New York Times* Page 3, 12:3, March 30, 1997). In 1993, the SEC reversed a long-standing policy by allowing Cracker Barrel Old Country Store to exclude a shareholder resolution which prohibited job discrimination on the basis of sexual orientation, arguing that diversity policy fell within the range of "ordinary business matters" which cannot by SEC rules be dictated by shareholder resolutions. Recently, shareholder interest groups have questioned the policy change. Our analysis suggests that shareholders may indeed have an interest in directing company policies towards diversity, even beyond the more widely cited issues of legal liability.

towards that type. One implication is that the entry of more women and minorities into the workforce will cause firms to shift their optimal bias towards these groups. Another implication is that in a many-level organization, diversity might fall as one goes up the hierarchy, since inequities in one level will be reinforced at the next higher level. This is consistent with the evidence cited in the introduction about the glass ceiling, as well as Bartlett’s (1997) description of diversity in the economics profession. In our model, it is even possible that a firm might have “affirmative action” at the entry level, but prefer majority types for high-level posts.

## 5 Markets for Entry-Level Workers

Thus far, we have abstracted from labor market considerations to focus on the evolution of diversity at a single optimizing firm with a stationary pool of entry-level workers. In this section, we provide a simple extension. The model includes two firms, a competitive labor market for entry-level employees, and a productivity benefit (unrelated to type or ability) from matching firms to workers. We then explore the effects of type-based mentoring on entry-level hiring policies and on wage differences between types. In particular, we examine conditions under which diversity of the entry level emerges in labor market equilibrium, as well as conditions under which the outcome is segregation, where some firms only hire type  $A$ ’s and other firms only hire type  $B$ ’s.<sup>25</sup>

We simplify the analysis by assuming linear initial ability functions and myopic behavior by firms. Index the firms by  $i = 1, 2$ . Let  $m_i^t$  denote the proportion of upper-level  $A$ -types at firm  $i$  in period  $t$  and let  $z_i^t$  denote the promotion policy at firm  $i$ . The mentoring differential at firm  $i$  is denoted  $\Delta_i = \mu(m_i^t) - \mu(1 - m_i^t)$ . Each firm hires a unit measure of lower-level employees at the beginning of each period. Let  $q_i^t$  be the proportion of lower level employees of type  $A$  at firm  $i$  in period  $t$ . In the labor market there is a unit measure of each type of employee, and the labor market must clear ( $q_1^t + q_2^t = 1$ ). Let  $w^t$  be the wage *differential* between  $A$ ’s and  $B$ ’s.

For both types, the initial ability for upper-level work at the population level is given by  $X(\theta) = 1 - \gamma\theta$  for  $\theta \in [0, 1]$ . Firms and workers observe initial ability *after* hiring and hence firms attract a random selection from each population. The productivity of a worker varies across firms. For half of the workforce, productivity is greater at firm 1 than at firm 2 by  $c \geq 0$ , while the other half is more productive at firm 2 by  $c$ . The quality of the firm-worker match is independent of type and initial ability and is observed before hiring (so that each firm hires its more productive workers first).

Our simple model incorporates an effect which should be present in more general

---

<sup>25</sup>In practice, we do sometimes observe segregated outcomes. Construction companies in some U.S. cities are dominated by a particular ethnic group. Some academic disciplines have large representations from particular nationalities, and gender integration is surprisingly heterogeneous across disciplines, even within the sciences.

settings. When the type  $A$  is in the majority in both firms ( $m_1^t, m_2^t \geq \frac{1}{2}$ ), type  $A$  is more productive in both firms, and the wage differential favors type  $A$ . However, when firms promote only a subset of their entry-level workers, there are decreasing returns to hiring entry-level workers of the majority type, since the probability that a given majority worker is promoted falls with the size of the majority pool.<sup>26</sup> Thus, firms may hire some workers of each type in equilibrium, even in the absence of firm-worker matching.

**Proposition 8** (i) Suppose  $m_1^t \geq \frac{1}{2}$  and  $r < 1$ . Then there exist values  $y \leq x \leq m_1^t$  such that for  $m_2^t \in [x, m_1^t]$ , entry levels are fully diverse ( $q_i^t = \frac{1}{2}$ ), for  $m_2^t \in (x, y)$ , entry levels are partially diverse ( $q_1^t \in (\frac{1}{2}, 1)$ ) and for  $m_2^t \leq y$ , entry levels are homogenous ( $q_1^t = 1$ ).

(ii) Entry levels are not fully diverse for all  $m_2^t \neq m_1^t$  (that is,  $x = m_1^t$ ), if and only if  $c = 0$ ; and entry levels are homogeneous for all  $m_2^t \neq m_1^t$  (that is,  $y = m_1^t$ ), if and only if  $c = 0$  and  $m_1^t = \frac{1}{2}$ .

(ii) At an interior equilibrium in which  $z_i^t \notin \{0, q_i^t, r\}$  and  $q_1^t \in (\frac{1}{2}, 1)$ :

$$w^* = \frac{\Delta_2 + \Delta_1}{\Delta_1^2 + \Delta_2^2} ((\Delta_1 - \Delta_2)c + r\Delta_1\Delta_2)$$

$$q_1^* = \frac{1}{2} + \gamma \frac{r(\Delta_1 - \Delta_2) - 2c}{\Delta_1^2 + \Delta_2^2}$$

We find that if the majority type differs for the two firms, then for  $c$  sufficiently small, the firms will specialize in choosing their entry-level workers. Workers will simply go to the firm where they are in the majority. However, when both firms have the same majority type, partial lower-level diversity is a possible outcome even when  $c = 0$ , since there are diminishing returns to entry-level workers of a given type. Further, with any positive amount of friction in the labor market ( $c > 0$ ), the entry level at each firm is fully diverse ( $q_i^t = 1/2$ ) if the firms are sufficiently similar. Although at full diversity, a more homogeneous firm always finds majority-type workers relatively more productive, any  $c > 0$  can overwhelm the difference between firms when they are sufficiently similar.

Under the assumption of myopic hiring, we can extend this one period analysis to consider the evolution of diversity at the two firms over time. As in the one firm model, upper level diversity at the firms follows a stochastic process with the state variable  $\{m_1^t, m_2^t\}$ .

**Corollary 9** (i) Suppose that  $m_s \in S^{UB}$ . Then  $\{m_s, m_s\}$  is a steady state in the two firm model; further, if  $c > 0$  and  $m_1^0$  and  $m_2^0$  are sufficiently close to  $m_s$ , the firms converge to  $\{m_s, m_s\}$ .

---

<sup>26</sup>Diversity would be even more likely as an outcome if, instead of a linear initial ability function, we allowed the initial ability function to be concave.

(ii) For  $c$  sufficiently small, segregated outcomes ( $m_1^t = 1 - m_2^t$ ) are steady states of the two firm model; and if initial firm asymmetry is sufficiently great ( $\{m_1^0, m_2^0\}$  sufficiently close to  $\{1, 0\}$ ) the firms converge to a segregated outcome.

The corollary shows that all the stable steady states of the single firm model with unbiased promotions are also symmetric steady states of the two firm model and that for  $c > 0$ , the firms will converge to these steady states if the firms start out sufficiently close. Of course, for  $c$  sufficiently small, full segregation is also a possible long-run outcome. Thus, the corollary provides some support for our single-firm model. Broadly, it implies that if periods are sufficiently long (so that firms are sufficiently short-sighted) and initial diversity levels are sufficiently similar, then any frictions in the labor market lead the firms to solve the optimization problem analyzed in our single-firm model. Thus, the insights of our single-firm model may be most applicable in industries where one type has historically dominated the industry, entry by new firms is difficult, and the existing firms differ in some way (such as location, culture, or specialty).

There are many ways to extend our two-firm model. For example, consider the case of far-sighted firms. As in the single-firm case, firms will bias their promotions. As firms shift their entry-level hiring to favor one type, the promotion bias should shift towards that type as well. Further, as  $\delta$  increases, the basin of attraction of the symmetric outcome should shrink, while that of the segregated outcome should increase. Other possible extensions include introducing labor supply decisions (which should lower the scarcity of majority-types since they receive a wage premium), and incorporating worker preferences about promotion prospects.<sup>27</sup>

## 6 Ex Ante Human Capital Investment

An important issue is the interaction between human capital investment decisions and firm promotion decisions when there is type-based mentoring. The greater the expected relative quality of the type  $A$  worker pool, the more type  $A$  workers a firm expects to promote in the future, and the more type  $A$  workers the firm wishes to promote today. This feeds back into the initial investment decisions of the workers. Anticipating a low probability of promotion even if their initial ability is high, minority-type workers in a given industry might choose to invest little in the specialized human capital required for upper-level work. At one extreme, this might lead to the clustering of a type in a few industries. In fact, such behavior has been observed historically, as is documented in the literature on “sex segregation” (Beller, 1984; Bergmann, 1989). Further, if one

---

<sup>27</sup> A benefit from promotion should work in favor of an interior equilibrium. Suppose  $A$ 's are in the majority at both firm 1 and 2 and all  $A$ 's are hired by firm 1. The promotion probability of  $A$ 's is higher at firm 2. Hence, the greater is the benefit to being promoted, the greater is the wage premium that firm 1 will have to pay to hold on to all of the  $A$ 's and hence the harder it will be to sustain full segregation as an equilibrium outcome.

type is the minority in most industries, that type might have lower incentives for human capital investment across the board.

Endogenous human capital investment could also have implications for public policy. Because of the interdependence of investment and promotion decisions, an economy may have multiple equilibria, one where  $B$  types have an inefficiently low level of investment, and another with high investment. In addition, one might see an inefficient clustering of minorities in only a few industries or occupations, when some of those minorities might have scarce talent that would be better matched in other industries. The government may wish to intervene to coordinate the economy on a high-investment equilibrium, perhaps by subsidizing education or training for type  $B$  workers in industries where they are under-represented, thereby inducing firms to make investments in promoting type  $B$  workers in anticipation of a talented worker pool.

## 7 Conclusion

This paper develops a novel perspective on discrimination and diversity. We begin with the view that firms possess a stock of upper-level employees. With type-based mentoring, the diversity of this stock matters, since it affects the flow of payoffs from newly promoted workers. The firm faces a tradeoff in its promotions between homogeneity, which maximizes the mentoring of a particular type, and diversity, which allows the firm to make use of scarce talent of workers of both types. The firm's promotion policies are thus biased, in the sense that firms sacrifice current profit in order to increase or decrease future diversity. We find that a firm's optimal bias may favor diversity or homogeneity, and that the attitude towards diversity depends on the curvature of the mentoring function, the strength of type-based mentoring, the scarcity of talent, the rate of retirement and the discount rate. Thus, our model provides some insight as to why some firms adopt policies of affirmative action, while others oppose such policies and even pass over minorities who have overcome the mentoring disadvantage. Our model can exhibit a glass ceiling, where the proportion of minorities in the upper level reaches a stable steady state which involves less diversity than at the lower level. Finally, we demonstrate that there may be multiple equilibrium levels of diversity. Then, full diversity may be the most profitable stable steady state for a firm, but it may not be optimal for a historically homogenous firm to sacrifice immediate profits to achieve full diversity without outside pressure.

Our model is predicated on several assumptions, notably that firms select from a lower-level workforce with talented workers of both types, and that firms capture some of the surplus from mentoring. Our model is most likely to be applicable for industries where firm-specific human capital is important and difficult to acquire through formal education, specialized or unusual skills are required, firms are long-lived, talents vary with the firm-worker match, internal labor markets are relatively more important than

external labor markets (at least for some parts of the hierarchy), and entry by new firms is uncommon. Examples that are especially appropriate include service firms and partnerships such as law, consulting, or accounting, investment banks, Fortune 500 companies, and academic departments. Our analysis explores how historical conditions affect long-run levels of diversity in such contexts.

Empirical evidence about the dynamics of diversity and its dependence on current levels of diversity would require data that tracks the diversity of organizations over time. Unfortunately, much of the publicly available data is collected at the level of individual workers, without identifying the employer, and thus there has not been much empirical research about diversity at the firm level. Further, it is difficult to identify a worker's place in the hierarchy of a firm from such data sources. Our study thus motivates the exploration of alternative data sources. For example, the data on partnership decisions by individual law firms is publicly available, and it would be especially appropriate for investigating the importance of mentoring effects.<sup>28</sup> More broadly, if firm-level data were available for a set of firms with similar hierarchical structures, it would be interesting to examine how a variety of individual outcomes (e.g. promotions, hiring, and turnover) depend on the type composition of the firm at different levels of the hierarchy.

Our model can be extended in a number of directions. One could further develop a theory of diversity in an economy by adding endogenous human capital acquisition by workers, as well as the entry and exit of firms. One could develop a more detailed micro model of mentoring (see Ragins (1997) for a discussion). For example, one could study the incentives to build mentoring relationships, or allow the level of mentoring received by a worker to depend on the ability of the mentor.

More broadly, there is the possibility to further develop a theory of the firm based on the accumulation of specialized assets over time (as in Prescott and Visscher (1980)). The business strategy literature highlights a number of assets which firms build up over time such as management talent, corporate culture and organizational capabilities (Barney, 1997, Dierickx and Cool, 1994). One can study various policies based on their effect on the accumulation of these assets, just as we study promotion policies based on their effect on the stock of upper-level employees.<sup>29</sup> Consider a firm's choice of which activities to engage in (e.g., which products to develop or customers to serve). Similarities in these activities across time give rise to organizational capabilities. A natural assumption is that a firm's capabilities affect the payoff from its current activities in a way similar to type-based mentoring: the more current activities are similar to past activities, the greater the current profit. This leads to a trade-off between breadth and depth of capability which appears surprisingly similar to the trade-off

---

<sup>28</sup>Unfortunately, the incidence of female or minority partners in law firms was very low during the 1980s, so only more recent data can potentially provide sufficient variation in diversity of partners.

<sup>29</sup>Rob and Zemsky (1997)—building on the general approach developed here—explore the effect of incentive policies on the evolution of a firm's stock of corporate culture.



between diversity and homogeneity that is the subject of this paper.

## 8 Appendix

**Lemma 10** *The value function defined by (1) is unique and continuous. The policy correspondence  $z^*$  is compact valued and upper hemi-continuous.  $V$  and  $|z^*(m) - r/2|$  are both symmetric about  $m = \frac{1}{2}$ .*

**Proof of Lemma 10:** In the notation of Stokey and Lucas (1989), our one period payoff function,  $F(m, m') = \pi(m, m' - (1 - r)m)$  is bounded and continuous; the function  $\Gamma(m) = [(1 - r)m, (1 - r)m + r]$ , which characterizes the feasible values of the state variable in the next period, is nonempty, compact valued and continuous; and the discount rate is in  $(0, 1)$ . Hence their Theorem 4.6 holds for our model and the value function is unique and continuous and the policy correspondence is u.h.c. The symmetry in  $V$  and  $|z^*(m) - r/2|$  follow from the symmetry of our model.  $\square$

**Proof of Proposition 1:** Recall that  $z^{UB}(m; r, \alpha, \gamma) = \arg \max_{z \in [0, r]} \pi(m, z; r, \alpha, \gamma)$ , and that  $\frac{\partial}{\partial z} \pi = x(z^{UB}; \gamma) - x(r - z^{UB}; \gamma) + \alpha [\hat{\mu}(m) - \hat{\mu}(1 - m)]$ . (i) By definition,  $x(z^{UB}; \gamma) - x(r - z^{UB}; \gamma)$  is nonincreasing in  $\gamma$  for  $z^{UB} \geq \frac{r}{2}$ , which holds if  $m \geq \frac{1}{2}$ . Thus,  $\frac{\partial^2}{\partial z \partial \gamma} \pi \leq 0$ , and  $z^{UB}$  must be nonincreasing in  $\gamma$ . But then,  $(r - z^{UB})/z^{UB}$  is nondecreasing in  $\gamma$ . (ii) Since  $\hat{\mu}(m) - \hat{\mu}(1 - m) > 0$  for  $m \geq \frac{1}{2}$ ,  $\frac{\partial^2}{\partial z \partial \alpha} \pi \geq 0$ , which implies that  $(r - z^{UB})/z^{UB}$  is nonincreasing in  $\alpha$ . (iii) If  $\frac{\partial}{\partial r} z^{UB} \cdot \frac{r}{z^{UB}} \leq 1$ , then  $(r - z^{UB})/z^{UB}$  is nondecreasing in  $r$ .  $\square$

**Proof of Proposition 2:** We proceed by induction on number of periods remaining in a firm's life. We know that when there is 1 period to go, the value function is equal to  $\pi^{UB}(m)$ , which is nonincreasing in  $m \geq \frac{1}{2}$  when SCV holds everywhere. Consider the problem with  $T$  periods to go, and assume that the value of the firm with  $T - 1$  periods to go,  $V^{(T-1)}(m)$ , is nonincreasing in  $m$ . The firm then solves

$$V^{(T)}(m) = \max_z \{ \pi(m, z) + \delta V^{(T-1)}((1 - r)m + z) \}$$

Observe that  $\frac{\partial^2}{\partial m \partial z} \pi(m, z) \geq 0$ . Since  $V^{(T-1)}(m)$  is nonincreasing, clearly  $z^{(T)}(m) \leq z^{UB}(m)$ , and so  $\frac{\partial}{\partial m} \pi(m, z^{(T)}) \leq \frac{\partial}{\partial m} \pi(m, z^{UB}) \leq 0$ . Using this and the envelope theorem,  $\frac{d}{dm} V^{(T)}(m) = \frac{\partial}{\partial m} \pi(m, z^{(T)}) + \delta \cdot (1 - r) V^{(T-1)'((1 - r)m + z^{(T)})} \leq 0$ . By induction, and since monotonicity is preserved by infinite sums, the infinite horizon value function is also nonincreasing in  $m \geq \frac{1}{2}$ . The argument is analogous for nowhere SCV.  $\square$

**Proof of Lemma 3:** (i) For  $M^{UB}(\cdot)$  nondecreasing, note that  $z^{UB}(m) = \arg \max_{z \in [0, r]} \pi(m, z)$  and  $\frac{\partial^2}{\partial m \partial z} \pi(m, z) > 0$ . Since the objective is differentiable, strictly concave, and strictly

supermodular,  $z^{UB}(m)$  must be unique everywhere, and strictly increasing when  $0 < z^{UB}(m) < r$ . (ii) Consider the following expression for  $M^*(m)$

$$M^*(m) = \arg \max_{m' \in \Gamma(m)} \pi(m, m' - (1-r)m) + \delta V(m'),$$

where  $\Gamma(m) = [(1-r)m, (1-r)m + r]$ . Note that the above objective is strictly supermodular in  $m$  and  $m'$  since

$$\frac{\partial^2 \pi(m, m' - (1-r)m)}{\partial m \partial m'} = \mu'(m) + \mu'(1-m) - x'(m' - (1-r)m) - x'(r - m' - (1-r)m) > 0. \quad (3)$$

Further,  $\Gamma(m)$  is nondecreasing in the strong set order (see Milgrom and Shannon, 1994). Hence,  $M^*(m)$  is nondecreasing and single-valued everywhere except for at upward jumps. Thus, every selection from  $M^*(m)$  is a continuous function but for upward jumps. (iii) Now, we establish convergence for both  $M^*$  and  $M^{UB}$ . Clearly  $M^{UB}(m > \frac{1}{2}) \geq \frac{1}{2}$ . Consider  $m' < \frac{1}{2}$  and  $m > \frac{1}{2}$ . Condition (3) and the symmetry of  $V$  implies that  $\pi(m, m' - (1-r)m) + \delta V(m') < \pi(m, 1 - m' - (1-r)m) + \delta V(1 - m')$ , which establishes that  $M^*(m > \frac{1}{2}) \geq \frac{1}{2}$  for  $m > \frac{1}{2}$ . By our constraint,  $M^*(m > \frac{1}{2}) \leq 1$ . Hence, following Milgrom and Roberts (1994), starting from  $m^0 > \frac{1}{2}$ , a fixed point of  $M^*$  and  $M^{UB}$  exists on  $[\frac{1}{2}, 1]$ .  $\square$

**Proof of Proposition 4:** (i)  $\frac{1}{2} \in S^{UB}$  if and only if  $z^{UB'}(\frac{1}{2}) \leq r$ . Using the implicit function theorem,  $z^{UB'}(\frac{1}{2}) = -\mu'(\frac{1}{2})/x'(\frac{r}{2})$ ; substituting gives the result. (ii) Consider  $x(\theta) = 1 - \theta$  and  $\mu(m) = rm$ . Then  $z^{UB}(m) = rm \forall m$  and  $S^{UB} = [0, 1]$ . It is then straightforward to perturb this  $x$  or  $\mu$  so that  $z^{UB}(m) \neq rm$  for  $m \in N$ , where  $N$  is any subset of  $(\frac{1}{2}, 1]$  which contains no isolated points.  $\square$

**Example 1:** Suppose  $x(\theta) = \kappa - \gamma\theta$ . Since  $M^{UB}(m) = (1-r)m + z^{UB}(m)$ , the curvature of  $M^{UB}$  follows that of  $z^{UB}$ . For  $z^{UB} < 1$ ,  $z^{UB}(m) = \frac{1}{2}r + (\mu(m) - \mu(1-m))/(2\gamma)$ . Thus,  $\frac{\partial^2 z^{UB}}{\partial m^2} = (\mu''(m) - \mu''(1-m))/(2\gamma)$ . Hence, if  $\mu'''(m) > 0$  for all  $m$ , then  $M^{UB}$  is concave over  $(\frac{1}{2}, 1]$  and there is at most one stable steady state in this interval; the steady state exists iff  $M'(\frac{1}{2}) > 1$ . If  $\mu'''(m) < 0$  for all  $m$ , then  $M^{UB}$  is convex over  $(\frac{1}{2}, 1]$  and there are at most two steady states in the interval, and only state  $m = 1$  can be stable.  $\square$

**Proof of Proposition 5:** (i) Follows by Milgrom and Roberts (1994), using our results on the sign of the bias from Proposition 2. (ii)  $z^{UB}(\frac{1}{2}) = \frac{r}{2}$ . Hence,  $z^*(\frac{1}{2}) = z^{UB}(\frac{1}{2}) + b(\frac{1}{2}) = \frac{r}{2}$  iff  $b(\frac{1}{2}) = 0$ , which is the case when SCV holds everywhere, but not when SCV holds nowhere.  $\square$

**Proof of Lemma 6:** Recall that Lemma 3 implies that  $M^*$  is single-valued almost everywhere, except at upward jumps. By definition,  $rm_s$  is one optimal promotion policy at state  $m_s$ . Suppose that there is another optimal promotion rule,  $z' > rm_s$ .

Then there must be an upward jump in  $M^*$  at  $m_s$ ; suppose that the jump is of size  $d$ . But this implies that for all  $d > \varepsilon > 0$ ,  $M^*(m_s + \varepsilon) > m_s + d > m_s + \varepsilon$ , contradicting the definition of a stable steady state.

Now consider an interior steady state  $m_s \notin \{0, \frac{1}{2}, 1\}$  and the associated steady state promotion policy  $z_s = rm_s$ . Since

$$z_s \in \arg \max_{z \in [0, r]} \pi(m_s, z) + \delta V(z + (1 - r)m_s)$$

it must be that  $\pi_2(m_s, rm_s) + \delta V'(m_s) = 0$  wherever this derivative exists. To solve for  $V'$ , let  $\vec{z} = (z^1, z^2, \dots)$ , and define

$$\tilde{V}(m, \vec{z}) = \pi(m, z^1) + \delta \pi(z^1 + (1 - r)m, z^2) + \dots$$

If  $m_s$  is a steady state, then  $V'(m_s) = \frac{d}{dm_s} \tilde{V}(m_s, \vec{z}_s)$  where  $\vec{z}_s = (rm_s, rm_s, \dots)$ . By the envelope theorem, we have that

$$\frac{d}{dm_s} \tilde{V}(m_s, \vec{z}_s) = \sum_{t=0}^{\infty} (1 - r)^t \delta^t \pi_1(m_s, rm_s) = \frac{\pi_1(m_s, rm_s)}{1 - (1 - r)\delta}.$$

Inserting this into the expression  $\pi_2(m_s, rm_s) + \delta V'(m_s) = 0$  gives the result.  $\square$

**Proof of Proposition 7:**(i) Applying the implicit function theorem, since by assumption  $m_s$  is optimal, it suffices to check that

$$\frac{\partial FOC_s(m; \gamma)}{\partial \gamma} = \frac{\partial [x(rm; \gamma) - x((1 - r)m; \gamma)]}{\partial \gamma} \geq 0$$

for  $m > \frac{1}{2}$ , which holds by definition. (ii) Following the arguments from the text, we apply the implicit function theorem. Using

$$\begin{aligned} \pi_1(m_s, rm_s) &= x(rm_s) - x(r(1 - m_s)) + \mu(m_s) - \mu(1 - m_s) \\ \pi_2(m_s, rm_s) &= m_s \mu'(m_s) - (1 - m_s) \mu'(1 - m_s) \end{aligned}$$

we have that

$$\frac{\partial FOC_s(m; \delta)}{\partial \delta} = \frac{r(m\mu'(m) - (1 - m)\mu'(1 - m))}{(1 - (1 - r)\delta)^2}.$$

Consider  $m > \frac{1}{2}$ . If  $\mu$  is globally SCV, then  $z^{UB}(m)\mu'(1 - m) > (r - z^{UB}(m))\mu'(m)$ . Then  $z^{UB}(m) > z^*(m) = rm$  implies that  $\partial FOC_s(m)/\partial \delta > 0$ . Conversely, if  $\mu$  is nowhere SCV,  $\partial FOC_s(m)/\partial \delta < 0$ . (iii) The first order conditions are given by

$$\frac{\partial}{\partial \alpha} FOC_s(m; \alpha) = \hat{\mu}(m) - \hat{\mu}(1 - m) + \frac{r\delta}{1 - (1 - r)\delta} (m\hat{\mu}'(m) - (1 - m)\hat{\mu}'(1 - m)).$$

Suppose that SCV holds nowhere. Observe that by Proposition 2 and optimality of the stable promotion policy, SCV nowhere implies that  $rm_s > z^{UB}(m_s)$ . But then, the second term in  $\frac{\partial}{\partial \alpha} FOC_s(m; \alpha)$  must be greater than  $\frac{\partial}{\partial m} \pi^{UB}(m)$ , which in turn is positive by SCV nowhere. For  $\delta$  small, recall that  $m_s \in S^{UB}$  satisfies  $z^{UB}(m_s) - rm_s = 0$ . Consider  $m_s \geq \frac{1}{2}$ . Then the diversity of  $S^{UB}$  is increasing (decreasing) in a parameter if  $z^{UB}(m_s) - rm_s$  is decreasing (increasing) in that parameter. It is straightforward to show that  $\partial z^{UB} / \partial \alpha > 0$ . (iv) In the linear case, we have

$$\frac{\partial}{\partial r} FOC_s(m; r) = -(2m - 1)\gamma + \frac{\delta(1 - \delta)}{(1 - (1 - r)\delta)^2} (m\mu'(m) - (1 - m)\mu'(1 - m)).$$

By Proposition 3 and optimality of the stable promotion policy, SCV everywhere implies that  $rm_s < z^{UB}(m_s)$ . But then, the second term in  $\frac{\partial}{\partial r} FOC_s(m; r)$  must be less than  $\frac{1}{r} \frac{\partial}{\partial m} \pi^{UB}(m)$ , which in turn is negative by SCV everywhere. For  $\delta$  small, we can consider whether  $\frac{\partial}{\partial r} (z_s^{UB} - rm_s) < 0$ , which holds if  $\partial z^{UB} / \partial r < m_s$ .  $\square$

**Proof of Proposition 8:** Suppose  $m_1 \geq \frac{1}{2}$  and  $r < 1$ . With random selection, the initial ability function for type  $A$  workers at firm  $i$  is  $x^A(\theta, q_i) = 1 - \gamma\theta/q_i$  while for type  $B$  workers it is  $x^B(\theta, q_i) = 1 - \gamma\theta/(1 - q_i)$ . The profit of firm  $i$  is then:

$$\begin{aligned} \Pi_i &= z_i \mu(m_i) + (1 - z_i) \mu(1 - m_i) \\ &\quad + \int_0^{z_i} x^A(\theta; q_i) d\theta + \int_0^{r - z_i} x^B(\theta; 1 - q_i) d\theta - q_i w - \left| q_i - \frac{1}{2} \right| c \end{aligned} \quad (4)$$

An equilibrium must be efficient. Because  $\frac{\partial}{\partial z_i} \Pi_i = \Delta_i - \gamma z_i / q_i + \gamma(r - z_i) / (1 - q_i)$ , a firm's optimal promotion policy (for a given a hiring policy) is

$$z_i^*(q_i) = \begin{cases} \min\{r, q_i\} & \text{if } \gamma r < q_i \Delta_i \\ 0 & \text{if } \gamma r < -(1 - q_i) \Delta_i \\ q_i(1 - q_i) \Delta_i / \gamma + q_i r & \text{otherwise} \end{cases} .$$

The marginal returns to hiring  $A$  types (for  $q_i \neq \frac{1}{2}$ ) are given by

$$\frac{\partial \Pi_i}{\partial q_i} = \begin{cases} \frac{\gamma z_i^2}{2q_i^2} - \frac{\gamma(r - z_i)^2}{2(1 - q_i)^2} - w + c(q_i), & z_i < q_i \\ \Delta_i & z_i > q_i \end{cases} \quad (5)$$

where  $c(q_i) = -c$  if  $q_i > 1/2$ , while  $c(q_i) = c$  if  $q_i < 1/2$ . Substituting in the optimal promotion policy  $z_i^*(q_i)$  into (5) we have that

$$\frac{\partial \Pi_i}{\partial q_i} - w + c(q_i) = \begin{cases} \gamma r^2 / (2q_i^2) & \text{if } \gamma r < q_i \Delta_i \text{ and } q_i > r, \\ -\gamma r^2 / (2(1 - q_i)^2) & \text{if } \gamma r < -(1 - q_i) \Delta_i \text{ and } (1 - q_i) > r, \\ (1/2 - q_i) \cdot \Delta_i^2 / \gamma + \Delta_i r & \text{if } \gamma r > \max\{q_i \Delta_i, -(1 - q_i) \Delta_i\}, \\ \Delta_i & \text{otherwise.} \end{cases}$$

and the marginal returns to hiring type  $A$ 's are nonincreasing so that

$$\frac{\partial \Pi_1}{\partial q_1} \Big|_{q_1=0} \geq \frac{\partial \Pi}{\partial q_1} \Big|_{q_1=\frac{1}{2}^-} \geq \frac{\partial \Pi_1}{\partial q_1} \Big|_{q_1=\frac{1}{2}^+} \geq \frac{\partial \Pi_2}{\partial q_2} \Big|_{q_1=1}. \quad (6)$$

We can use this to prove parts (i) and (ii). Note that as  $m_2 \rightarrow m_1$ ,  $\partial \Pi_2 / \partial q_2 \rightarrow \partial \Pi_1 / \partial q_1$  and hence there exists values  $y \leq x \leq m_1^t$  such that for  $m_2^t \in [x, m_1^t]$ ,  $q_i^t = \frac{1}{2}$ , for  $m_2^t \in (x, y)$ ,  $q_1^t \in (\frac{1}{2}, 1)$  and for  $m_2^t < y$ ,  $q_1^t = 1$ . Further, the second inequality in (6) holds with equality and hence  $x = m_1$  iff  $c = 0$ . The first term in (6) is equal to the last term and hence  $y = m_1$  iff  $c = 0$  and  $\Delta_1 = 0$  ( $\Leftrightarrow m_1 = \frac{1}{2}$ ).

To prove (iii), denote equilibrium quantities by  $z_i^*$  and  $q_i^*$ . We now characterize an interior equilibrium where  $z_i^* \notin \{0, q_i^*, r\}$  for  $i = 1, 2$  and  $q_1^* \in (\frac{1}{2}, 1)$  ( $\Leftrightarrow \Delta_1 \geq \Delta_2$ ). Solving  $\partial \Pi_i / \partial q_i = 0$  for  $q_i$  in this case yields  $q_1(w) = 1/2 + r\gamma/\Delta_1 - (w+c)\gamma/\Delta_1^2$ , and  $q_2(w) = 1/2 + r\gamma/\Delta_2 - (w-c)\gamma/\Delta_2^2$ . Imposing market clearing,  $q_1(w) + q_2(w) = 1$ , we find the expressions in the proposition.  $\square$

**Proof of Corollary 9:** (i) Suppose  $m_1 = m_2 \geq \frac{1}{2}$ . Then  $q_1^* = q_2^* = \frac{1}{2}$  (unless  $m_1 = m_2 = \frac{1}{2}$  and  $c = 0$ , in which case any  $q_1^* \in [0, 1]$  is an equilibrium). Further,  $z_1^* = z_2^*$  and in the following period  $m_1 = m_2$  as well. The dynamics are then exactly as in the one firm model. If  $c > 0$ ,  $q_1^* = q_2^* = \frac{1}{2}$  for  $m_1, m_2$  sufficiently close to  $m_s$ , and thus  $m_s \in S^{UB}$  implies that the firms converge to  $m_1 = m_2 = m_s$ . Suppose now that  $m_1$  and  $m_2$  are arbitrarily close to  $m_s$  but that  $m_1 > m_2$ . Again,  $q_1^* = q_2^* = \frac{1}{2}$ . Since  $m_s$  is stable with myopic promotions and  $q_i^* = \frac{1}{2}$ , the firms converge to this outcome. (ii) Suppose  $m_1 > \frac{1}{2} > m_2$ . Then for  $c$  sufficiently small,  $\partial \Pi_1 / \partial q_1 + w > 0 > \partial \Pi_2 / \partial q_2 + w$  and hence  $q_1^* = 1$ , which implies that  $z_1^* = r$  and  $z_2^* = 0$ . Hence, the firms converge to  $m_1 = 1, m_2 = 0$ .  $\square$

## 9 References

- Arrow, Kenneth (1973). "The Theory of Discrimination," in O. Ashenfelter and A. Rees (eds.) *Discrimination in Labor Markets*, Princeton University Press.
- Barney, Jay B. (1997). *Gaining and Sustaining Competitive Advantage*, New York: Addison Wesley.
- Barlett, Robin, 1997, "Report of the Committee on the Status of Women in the Economics Profession," *American Economic Review* 87(2): 506-510.
- Bergmann, Barbara (1989). "Does the Market for Women's Labor Need Fixing?" *Journal of Economic Perspectives*, 3, Winter, 43-60.
- Beller, Andrea (1984). "Trends in Occupational Segregation by Sex and Race, 1960-1981," in *Sex Segregation in the Workplace*, edited by Barbara F. Reskin, Washington, DC: National Academy.
- Carmichael, Lorne (1988). "Incentives in Academics: Why Is There Tenure?" *Journal of Political Economy* 96 (3): 453-472.
- Coate, Stephen and Loury, Glenn (1993). "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review* December, 1220-1240.
- Cain, Glen (1986). "The Economic Analysis of Labor Market Discrimination: A Survey," in *The Handbook of Labor Economics*, edited by Orley Ashenfelter and Richard Layard, New York: Elsevier, Science.
- Cornell, Bradford and Ivo Welch (1996). "Culture, Information and Screening Discrimination," *Journal of Political Economy* 104 (3):542-572.
- Dierickx, Ingemar and Karel Cool (1994), "Competitive Strategy, Asset Accumulation and Firm Performance" in *Strategic Groups and Strategic Moves* edited by H. Daems and H. Thomas, Oxford: Pergamon Press.
- Dreher, George F., and Ash, Ronald A. (1990), "A comparative study of mentoring among men and women in managerial, professional, and technical positions." *Journal of Applied Psychology* 75 (October) pp. 539-46.
- Dreher, George F., and Cox, Taylor H. (1996), "Race, gender, and opportunity: a study of compensation attainment and the establishment of mentoring relationships." *Journal of Applied Psychology* 81 (June) pp. 297-308.
- Gilson, Ronald and Mnookin, Robert (1989). "Coming of Age in a Corporate law Firm: The Economics of Associate Career Patterns," *Stanford Law Review* 41 (February): 567-595.
- Greif, Avner (1994), "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102 (5) October, pp. 912-950.
- Greif, Avner (1993), "Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition," *American Economic Review* 83 (3) June, pp. 525-48.

- Ibarra, Herminia (1992), "Homophily and differential returns: sex differences in network structure and access in an advertising firm." *Administrative Science Quarterly* 37 (September) pp. 422-47.
- Kanter, Rosabeth (1977). *Men and Women of the Corporation*, New York: Basic Books, Inc.
- Laband, David N., and Lentz, Bernard F. (1994), "Workplace mentoring in the legal profession," *Southern Economic Journal* 61 (January) pp. 783-802.
- Milgrom, Paul and Roberts, John (1994). "Comparing Equilibria," *American Economic Review*, June, 441-459.
- Milgrom, Paul and Shannon, Christina (1994). "Monotone Comparative Statics," *Econometrica*, 60 (1), p. 157-80.
- Morrison, Ann, and Von Glinow, Mary Ann (1990). "Women and Minorities in Management." *American Psychologist*, February, 200-208.
- Noe, Raymond (1988). "Women and Mentoring: A Review and Research Agenda." *Academy of Management Review*, 13 (1): 65-78.
- Orpen, Christopher (1995). "The effects of mentoring on employees' career success." *The Journal of Social Psychology* 135 (October) pp. 667-8.
- Prescott Edward and M. Visscher (1980). "Organizational Capital." *The Journal of Political Economy* 88: pp. 446-461.
- Riley, Sandra and Wrench, David (1985), "Mentoring among women lawyers." *Journal of Applied Social Psychology* 15 (4): pp. 374-86.
- Ragins, Belle Rose (1997), "Diversified Mentoring Relationships in Organizations." *Academy of Management Review* 22 (2): pp. 482-521.
- Rosen, Asa (1997). "An Equilibrium Search-Matching Model of Discrimination." *European Economic Review*, 41: 1589-1613.
- Rothstein, Donna (1997). "Supervisor Gender, Race, and Ethnicity and the Labor Market Outcomes of Young Workers." Mimeo, U.S. Bureau of Labor Statistics.
- Spurr, Stephen, 1990, "Sex Discrimination in the Legal Profession: A Study of Promotion." *Industrial and Labor Relations Review* 43 (April): 406-417.
- Stokey, Nancy, and Lucas, Robert (1989). *Recursive Methods in Economic Dynamics*, Cambridge: Harvard University Press.