Organizational Spirals and Spontaneous Affirmative Action

Paul-Henri Moisson[†,*] Jean Tirole[†,**]

May 14, 2020

Abstract

The paper studies discriminatory hiring by an organization whose members' cooptation decisions are driven by two motives: quality and homophily. Small variations in the initial quality or diversity of the organization may lead to virtuous or vicious spirals. In order to avoid depriving itself of its talent pool, an organization voluntarily engages in affirmative action if and only if its quality-diversity mix is currently unattractive yet curable.

Keywords: Organizations, affirmative action, virtuous and vicious spirals.

JEL Codes: D7, D02, M5.

[†] The authors are grateful to participants at numerous seminars and conferences. Both authors gratefully acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 669217 – ERC MARK-LIM), and from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program (grant ANR-17-EURE-0010).

[*] Toulouse School of Economics (TSE), paulhenri.moisson@tse-fr.eu.

[**] Toulouse School of Economics (TSE) and Institute for Advanced Study in Toulouse (IAST), jean.tirole@tse-fr.eu.

1 Introduction

The paper analyzes hiring by a large organization whose members' cooptation decisions are driven by two motives: quality and homophily. It shows that small exogenous variations in the initial quality or diversity of the organization (e.g. due to staffing disruptions, technological shocks, globalization. . .) may lead to virtuous or vicious spirals and markedly different steady states. For, talented minorities may refuse to join an organization that lacks diversity (all the more so when its average quality is low). The higher the initial diversity and the higher the initial quality, the more likely is the organization to converge to a high-quality, high-diversity steady state. If the quality and diversity are initially low, the organization does not attract talented minorities. For an even worse quality-diversity mix, it even stops attracting talented majority candidates. There is a region, though, over which, in order to avoid depriving itself of its talent pool, an organization voluntarily engages in affirmative action, picking minority candidates over at-least-equally talented majority candidates.

Our model has continuous time. At each instant, there is a flow of departing and incoming members; the latter are selected through majority voting by existing members. There are two groups (differing in gender, religion, ethnicity, background, politics, scientific field or approach, values, etc.). The hiring pool includes candidates of both groups, but only a subset of those are talented: talent is in short supply. A talented hire brings extra utility (knowledge, prestige, budgets, etc.) to all other members, while homophily benefits accrue only to members of the same group. To avoid trivial dynamics in which all recruiting is in-group, we assume that quality benefits exceed homophily ones. Members are forward looking, and so are potential hires. Talented candidates have higher outside options than untalented ones. They can be attracted only if the present discounted value of the payoff in the organization exceeds the outside option; this comparison requires anticipating on the evolution of the organization.

We thus analyze a dynamic game whose players are far-sighted incumbent members, minority and majority candidates, and their future counterparts. We look for a Markov perfect equilibrium, guess strategies in the state space and verify that these strategies are indeed optimal for all players. As announced, its dynamics involve both affirmative action and virtuous and vicious spirals, depending on the starting point.

Related literature. We will not review the large literature on hiring discrimination[1]. Our paper, to the best of our knowledge, is the first to show that voluntary affirmative action may result from a concern for not being sufficiently attractive to minority employees.

In its emphasis on virtuous and vicious spirals, the paper is most closely related to Board et

[1] See e.g. the literature reviews in Board et al (2019), Cai et al (2018) – another paper stressing the mix of talent and homogamy concerns in hiring, but with a different emphasis –, and Moisson-Tirole (2020).

al (2019). The latter contribution derives rich dynamics from the reasonable assumption that talented people are better at identifying new talents. It shows that high-skill firms post high wages, screen applicants first, extract talent from the applicant pool, and exert a negative compositional externality on low-skill, low-wage firms. Consequently, talent is a source of sustainable competitive advantage. Their virtuous and vicious spirals have a different origin from ours: a stronger organization makes better hiring choices. Their analysis focuses on the vertical (quality) dimension. In our paper, talented minority (and perhaps also talented majority) candidates turn down an organization that lacks diversity and/or talent.

2 Model

Payoffs in organization and outside options

Like in Cai et al (2018) and Moisson-Tirole (2020), organizational members have a two-dimensional type; they differ in their talent and the group they belong to. Time is continuous rather than discrete. The horizon is infinite: $t \in (-\infty, +\infty)$. The organization has a unit mass of members. Each individual has a two-dimensional type. The vertical type captures ability or talent and takes one of two possible values $\{0, \tilde{s}\}$, where $\tilde{s}dt > 0$ is the incremental flow contribution of a talented individual to each other member's payoff. The horizontal type stands for race/gender/tastes/opinions and can take two values $\{A, B\}$. A member of a given horizontal type exerts flow externality $\tilde{b}dt > 0$ (where $\tilde{b} > 0$) on members of the same type, but not on members of the opposite type, and this regardless of their talent.

The organization is characterized by a state $X \equiv \{M, S\}$: $M \in [1/2, 1]$ is the majority's size (where majority and minority are defined with respect to horizontal types) and S the fraction of talented members (so that the current quality of the organization is equal to $S\tilde{s}$), the flow payoffs of a majority and a minority member are, respectively

$$[S\tilde{s} + M\tilde{b}]dt \quad \text{and} \quad [S\tilde{s} + (1-M)\tilde{b}]dt.$$

Between times t and $t + dt$, a fraction χdt of incumbent members exits, and χdt new members are coopted. During this interval of time, there is a large number (an excess supply) of untalented candidates of each group, as well as $x\chi dt$ talented candidates from each group, where $x < 1/2$. Candidates can enter the organization when they arrive (and only then) and have a death rate equal to χ inside or outside the organization (their discount rate is $r + \chi$, where r is the pure rate of time preference).

Talented and untalented candidates differ in their outside option. Talented candidates obtain flow payoff $\tilde{u}dt$ outside the organization, untalented ones a zero flow payoff. So, a talented candidate accepts an offer if and only if their utility, i.e. the discounted sum of their flow payoffs, is greater than or equal to that in the outside option $\tilde{u}/(r+\chi)$, while an untalented candidate always accepts an offer. We will first look for equilibria in which only the talented minority candidates' participation constraint is binding.

While an agent's effective discount factor is $r+\chi$, the (quality or homophily) benefits that a new member brings to a given current member must be discounted by $r+2\chi$ since the flow probability that either the current member or the new member exits the organization is $2\chi dt$. We accordingly define the expected intertemporal utilities: $s \equiv \tilde{s}/(r+2\chi)$, $b \equiv \tilde{b}/(r+2\chi)$. To keep surpluses comparable, we similarly define $u \equiv \tilde{u}/(r+2\chi)$.

Hiring and acceptance strategies

Candidates' participation decisions are intertemporal strategic complements: a candidate is more willing to join the organization today if she knows that talented candidates will be more prone to join it in the future. For the sake of simplicity, we shall focus on equilibria in which there are no intertemporal coordination failures among talented candidates of the same group or different groups.

Let σ_1 (resp. σ_2) denote the (state-contingent) fraction of talented candidates of the majority (resp. minority) who are selected by the majority – later on, we will note that in equilibrium $\sigma_1 = \sigma_2 = 1$. Let σ_0 denote the fraction of remaining slots $\big(1-x(\sigma_1+\sigma_2)\big)$ that are allocated to untalented majority candidates. Thus, $\sigma_0 < 1$ indicates some voluntary affirmative action (the majority selects untalented out-group candidates over equally untalented in-group ones).[2]

Note that in large organizations facing a symmetric talent pool, the majority is freed from the vagaries of a random pool of candidates, and never faces a tradeoff between sacrificing quality and losing control. We therefore look for an equilibrium in which the majority solves an optimal control problem, without having to worry about the possibility of losing control. The majority's program writes as

$$\max_{\sigma_0,\sigma_1,\sigma_2} \int_0^{+\infty} e^{-(r+\chi)t} \Big[S_t \tilde{s} + M_t \tilde{b} \Big] dt$$

subject to the participation constraints of talented candidates, and the induced dynamics of S_t and M_t.

We make two assumptions, that together will guarantee the existence of a steady state at which all talented candidates are willing to join the organization (which does not mean that they will do so in other states).

Assumption 1. *(The organization may attract minority talents) Under parity and attractive-*

[2] This of course can be viewed as a weak form of affirmative action. However, the same dynamics would hold if the in-group untalented candidates were slightly more productive than their out-group counterparts.

ness for talented candidates, talented members receive a positive net surplus:

$$u < 2xs + \frac{b}{2}. \tag{1}$$

Assumption 2. *(Minority talents' outside option constrains the majority) A talented minority candidate does not want to join a strongly homogenous organization, even a high-quality one: A steady-state absence of affirmative action (namely, $\sigma_0 = \sigma_1 = \sigma_2 = 1$) is bound to put off talented minority candidates:*

$$2xs + xb < u. \tag{2}$$

Let

$$\frac{1}{2} < M^* \equiv \frac{2xs + b - u}{b} < 1.$$

3 Equilibrium dynamics

When talented minority and majority candidates accept to become members (regions 1 and 2 below), the flow-quality dynamics are given by

$$\frac{dS}{dt} = \chi[-S + 2x]$$

These dynamics are autonomous and converge monotonically to $S^* \equiv 2x$. Those for the majority size by contrast depend on the majority's strategy and therefore on the state $\{S_t, M_t\}$ of the organization:

$$\frac{dM}{dt} = \chi[-M + x + (1-2x)\sigma_0(S, M)]$$

The equilibrium exhibits (at most) four regions when the talented majority candidate's outside option is not binding:

- Region 1 (*standard favoritism*): when $Ss + (1-M)b > u$ (talented minority members enjoy a flow surplus in the organization), the majority favors its own candidates in the untalented group ($\sigma_0 = 1$):

$$\frac{dM}{dt} = \chi[-M + (1-x)]$$

- Region 2 (*mild affirmative action to keep talented minority candidates on board*): when $Ss + (1-M)b = u$, the majority selects candidates so as to maintain minority indifference between being in the organization or outside the organization:

$$s\frac{dS}{dt} = b\frac{dM}{dt} \qquad \Longleftrightarrow \qquad \sigma_0 = \frac{2xs + (1-x)b - u}{(1-2x)b} \equiv \sigma_0^*$$

Assumption 1 implies that $\sigma_0^* > 0$, while Assumption 2 implies that $\sigma_0^* < 1$. Whenever the organization reaches region 2, it monotonically converges to the steady state (S^*, M^*), which lies in region 2.[3]

- Region 3 (*strong affirmative action to make the organization attractive to the minority again*): when $M \leq \phi(S)$ (for some increasing ϕ satisfying $Ss + (1-\phi(S))b < u$), the majority selects $\sigma_0 = 0$. Talented minority candidates turn down offers (they receive negative net utility until region 2 is reached and zero net utility thereafter). Dynamics are given by

$$\frac{dS}{dt} = \chi[-S + x], \qquad \text{and} \qquad \frac{dM}{dt} = \chi[-M + x]$$

- Region 4 (*giving up on minority candidates*): the majority selects only majority candidates, as the "investment cost" to make the organization sufficiently attractive to talented minority candidates is too large. Dynamics are described by

$$\frac{dS}{dt} = \chi[-\tilde{S} + x], \qquad \text{and} \qquad \frac{dM}{dt} = \chi[-M + 1]$$

Hence, whenever the organization reaches region 4, it monotonically converges to the steady state $(x, 1)$, which lies in the interior of the region.

Being willing to do what it takes to attract talented minority members requires that the quality payoff s be sufficiently high. We therefore henceforth assume:

Assumption 3. *(Affirmative action may be attractive) The majority's flow payoff from newcomers is higher in the high-quality steady state than in the low-quality one:*

$$2xs + xb + (1-2x)\sigma_0^* b > xs + b \qquad \Longleftrightarrow \qquad 3xs > u. \tag{3}$$

Let us check the optimality of talented minority members' joining decision (they join the organization in regions 1 and 2, but not in the other regions). We can distinguish two groups of regions: $\mathcal{R}^+$ is composed of regions 1 and 2, in which talented minority members enjoy either a strictly positive instantaneous net surplus (region 1) or a zero net surplus (region 2).

[3] By contrast, if $u \in [2xs, 2xs + xb)$, region 2 would never be reached and the steady state would be given by $(S^*, 1-x)$ and be interior to Region 1.

[4] Figure 1 for complete generality allows s to exceed $2x$; for instance, there might have been a more favorable supply of talent prior to date 0.

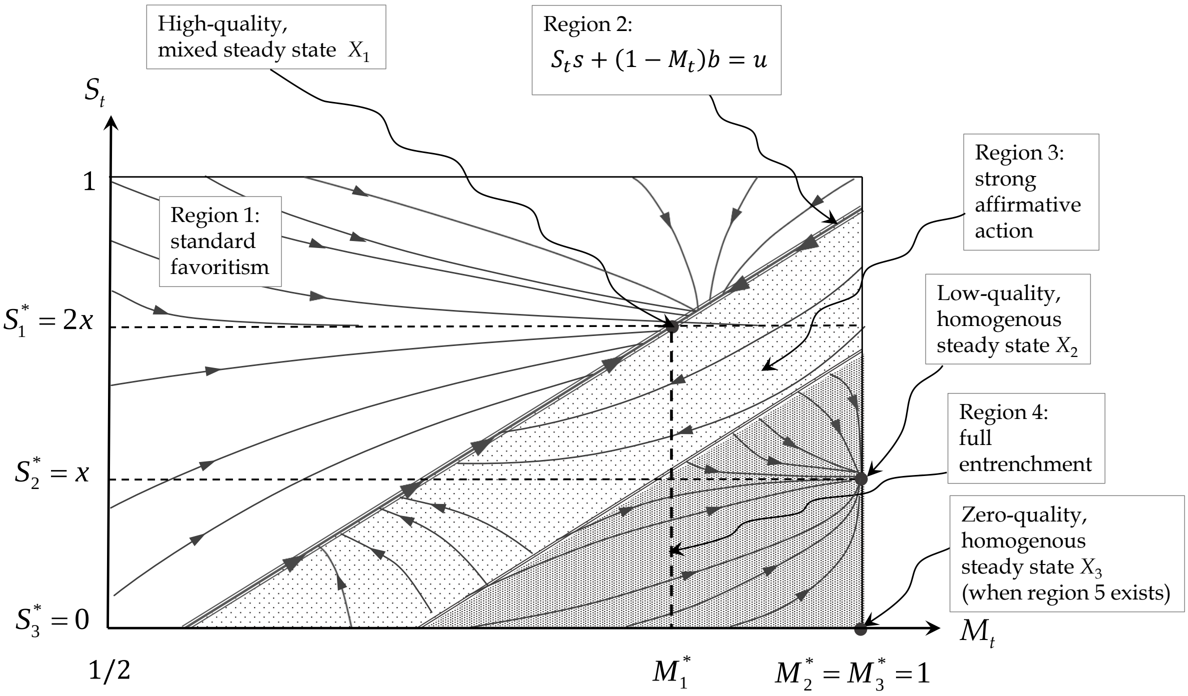

Figure 1: Phase diagram of the organization's current quality S and majority size M when $u \in [2xs + xb, 2xs + b/2]$ and $u < \max(xs + (1-x)b, b/2, 3xs)$, so region 5 does not exist.[4]

Because $\mathcal{R}^+$ is absorbing, a talented minority member enjoys a non-negative net surplus in each future period, implying the optimality of acceptance. $\mathcal{R}^-$ is composed of regions 3 and 4 (and possibly 5, see below), which all deliver a strictly negative instantaneous net surplus; even if organizational dynamics converge to absorbing region 2, which gives them a zero net surplus, their net utility of joining overall is strictly negative.

The Online Appendix shows that region 3 is non-empty if and only if the following condition holds:

Assumption 4. *(Affirmative action can lure back talented minority candidates).*[5,6] *New hires under full affirmative action bring positive net surplus to the minority and therefore improve the organization's attractiveness relative to legacy outside regions 1 and 2 (which yields mi-*

[5] With full affirmative action ($\sigma_0 = 0$), the dynamics for $(Ss - Mb)$ write as $\frac{d}{dt}(Ss - Mb) = \chi\big[-Ss + Mb + x(s-b)\big]$. Hence the minority will ever be willing to join the organization only if $\lim_{t \to +\infty}(S_t s + (1 - M_t)b) > u$, i.e. $xs + (1-x)b - u > 0$.

[6] The assumption that $u \in [2xs + xb, 2xs + b/2]$ combined with (4) implies in particular that $x < 1/3$.

nority members a negative net surplus)

$$xs + (1 - x)b - u > 0 \tag{4}$$

If (4) holds, then whenever $u \leq b/2$, region 3 is given by the set of states (M_0, S_0) such that *the majority's sacrifice is worth the trouble starting at* (M_0, S_0): for $S_0 s - M_0 b < u - b$,[7]

$$\int_0^T e^{-(r+\chi)t}(1-x)b\big[1 - e^{-\chi t}\big]dt + \int_T^{\infty} e^{-(r+\chi)t}(1-x)b\big[1 - e^{-\chi T}\big]e^{-\chi(t-T)}dt$$
$$\leq \int_T^{+\infty} e^{-(r+\chi)t}(3xs - u)\big[1 - e^{-\chi(t-T)}\big]dt \tag{5}$$

where T is given by

$$T \equiv \frac{1}{\chi}\ln\left[\frac{M_0 b - S_0 s + x(s-b)}{xs + (1-x)b - u}\right] \geq 0$$

Condition (5) may thus be rewritten in terms of initial attractiveness for the minority: $\underline{u} < S_0 s + (1 - M_0)b < u$ for some $\underline{u} > 0$.

Failing to attract talented majority members. The outside option u is non-binding for talented majority members if the organization is attractive to talented minority members (regions 1 and 2): let us thus consider their participation constraint when $Ss + (1-M)b < u$. A fifth region may exist in which the organization fails to attract any talented candidate, and subsequently converges towards homogeneity and zero-quality ($M_3^* = 0$, $S_3^* = 0$). A necessary condition for this "region 5" to be non-empty is $u > b/2$. Then, if this inequality holds, region 5 is in particular non-empty for χ sufficiently close to 0 and (with additional conditions) for χ sufficiently high, i.e. if turnover is sufficiently low or sufficiently high. The intuition underlying this result is that when turnover is too low, the organization fails to renew its composition fast enough, whereas when turnout is too high, members are likely to quit the organization before they could reap the benefits of quality improvement. We provide more details in the Online Appendix, and summarize the key messages in Proposition 1.

Proposition 1. ***(Voluntary affirmative action and virtuous/vicious spirals)*** *Under Assumptions 1 and 2, there exists an MPE satisfying:*

(i) Path uniqueness and steady states: There exist at least two steady states: $(M_1^*, S_1^*) =$

[7] As the LHS in (5) is strictly positive for $T = 0$ (from assumption (3)), condition (5) holds by continuity for T in a neighbourhood of 0, i.e. for any couple (M_0, S_0) in a neighbourhood of the line $\{(M,S) \mid Mb - Ss = b - u\}$. Furthermore, since the LHS in (5) strictly increases with T while the RHS strictly decreases with T, if condition (5) is satisfied by a couple (M_0, S_0), then it holds by continuity for any initial state (M_0', S_0) such that $M_0' \leq M_0$. Lastly, by monotonicity, there exists a unique T such that (5) holds with equality. Hence the boundary between Regions 3 and 4 is an (increasing) line in the plane (M, S).

$(x + (1 - 2x)\sigma_0^*, 2x)$ and $(M_2^*, S_2^*) = (1, x)$. A third steady state, $(M_3^*, S_3^*) = (1, 0)$, may exist. Both majority and minority members rank the steady state (M_1^*, S_1^*) first, (M_2^*, S_2^*) second and (M_3^*, S_3^*) third. There are at most 5 regions in the state space $\{(M, S) \mid M \in [1/2, 1], S \in [0, 1]\}$. Starting from an initial state[8]*, there exists a unique equilibrium; the organization converges to the region's steady state (which may or may not be interior to the region).*

(ii) *Path dependence: A lower initial quality S_0 generates a lower steady-state quality for the organization, and a lower steady-state utility for both the majority and the minority members (if any)*[9]*. Similarly, absent an outside option for talented majority candidates or if $u - b/2 < xs$, a larger initial majority size M_0 has a long-run impact pointing in the same direction as a lower initial quality S_0. By contrast, if talented majority candidates also have the outside option u and if $u > b$ (resp. $u - b/2 < xs$), a larger initial majority size M_0 may enhance the organization's steady-state quality if it allows the organization to attract talented majority candidates and converge towards the steady states $(x + (1 - 2x)\sigma_0^*, 2x)$ or $(1, x)$ instead of $(1, 0)$ (i.e. move out of region 5 into regions 3 or 4).*[10]

(iii) *Voluntary affirmative action: Under Assumptions 3 and 4, there exists a range of initial states (region 3) in which the majority engages in voluntary affirmative action in order to get talented minority candidates back on board in the future (and subsequently weakly reducing affirmative action).*

(iv) *Vicious spirals: The union of regions 4 and 5 is absorbing. The organization either converges towards (M_2^*, S_2^*) or (M_3^*, S_3^*) (the latter if and only if the initial state lies in region 5 and $u > b$).*[11]

Remark 1: Quits. If the organization's talented current members also have access to the same outside option as outside candidates, then the organization may lose all its talented minority members at once. This would put an additional constraint on the profitability of engaging in affirmative action, thus reducing the size of region 3. Upon losing all its talented minority members, the organization goes from a state (M_{t^-}, S_{t^-}) to a state (M_{t^+}, S_{t^+}) where M_{t^+} and

[8] Except along the non-generic line $M = \phi(S)$, and the boundary between regions 4 and 5, if any.

[9] As S^*, $M^*b + S^*s$ and $(1 - M^*)b + S^*s$ weakly increase with S_0.

[10] In the steady state of region 2, the relative talent of majority vs. minority members is strictly below 1 and higher than in the "objective state" of region 1. The talent ratio of majority vs. minority members decreases over region 5. It may increase or decrease over regions 1, 2, 3 and 4 depending on the initial composition of the organization.

[11] Moreover, it can be shown that in some cases at least, in the long run, as the quality decreases from its initial level S_0, talented minority candidates reject the organization's offers before talented majority candidates do.

S_{t+} depend on the initial distribution of quality *within each group.*

Remark 2. Deterministic, symmetric models of large organizations imply that control is no longer an issue: Meritocracy allows the majority to keep control. This feature is inconsequential for the investigations of various dynamics under entrenchment, as is the case here. Furthermore, one can introduce control concerns in the large-organization model by adding persistent shocks. For example, members may not always vote and the relative absenteeism of the majority vs. minority members might follow a Brownian motion. Another (but asymmetric) environment in which control is a concern for the majority arises when there are more talented B-candidates than talented A-candidates: $x_A < x_B$ and group A initially has the majority.[12]

4 Competition for talent

Competition among organizations is a rich object of study, which we leave for future work. We here content ourselves with a partial result highlighting the analogy with endogenous candidacies; the key difference with the case of endogenous candidacies is that outside options are strategically determined by rival platforms.

Suppose there are two large organizations $j = 1, 2$. As earlier, in each time interval $[t, t{+}dt]$, there is a mass χdt of departing members in each organization, and a mass $\chi x dt$ of talented candidates of each group, together with an unlimited supply of untalented candidates of either type. Each organization ranks-order candidates; when confronted with multiple acceptances, candidates pick their preferred organization and the market clears by moving down the organizations' pecking order. There are now five state variables: $\{M_j(t), S_j(t)\}_{j\in\{1,2\}}$ together with whether the majoritarian groups are the same in the two organizations. When considering an organization's offer, candidates have no other outside option than the other organization's

[12] If $x_B > 1/2$ and if there is a benefit from control ($x_A + x_B < 1$), then an A-majority may face a tradeoff between engaging in affirmative action in order to attract talented B-candidates, and retaining control. We provide an illustration in the case where there are no outside options. Assume for instance that between times t and $t + dt$, there are $x_A\chi dt$ (resp. $x_B\chi dt$) talented candidates from group A (resp. B), where $x_A < 1/2 < x_B$ and $x_A + x_B \leq 1$ (as well as a large number of untalented candidates of each group, as before). Whenever control is not at stake, the majority still favours talented out-group candidates over untalented in-group ones. Yet consider an A-majority with size 1/2. The majority may then either relinquish control, in which case the flow quality (resp. homophily) payoff of A-members will converge toward $(x_A + x_B)\tilde{s}$ (resp. $x_A\tilde{b}$), or keep control, in which case the flow quality (resp. homophily) payoff of A-members will converge toward $(x_A+1/2)\tilde{s}$ (resp. $\tilde{b}/2$). Hence an A-majority with size 1/2 chooses to relinquish control if and only if

$$\int_0^\infty e^{-(r+\chi)t}\Big[\big[S_0-(x_A+x_B)\big]\tilde{s}e^{-\chi t}+(x_A+x_B)\tilde{s}+\big[1/2-x_A\big]\tilde{b}e^{-\chi t}+x_A\tilde{b}\Big]dt$$
$$\geq \int_0^\infty e^{-(r+\chi)t}\Big[\big[S_0-(x_A+1/2)\big]\tilde{s}e^{-\chi t}+(x_A+1/2)\tilde{s}+\tilde{b}/2\Big]dt$$

i.e. if and only if $s/b \geq (1/2 - x_A)\big/(x_B - 1/2)$.

(potential) offer.

We say that an equilibrium is "group-coalition proof" if talented candidates of a group cannot deviate from their acceptance strategies and all be better off; the equilibrium is "population-coalition proof" if talented candidates of both groups cannot deviate from their acceptance strategies and all be better off[13]. We focus on "increasing-dominance equilibria", i.e. equilibria such that (a) both organizations recruit all talented candidates willing to join the organization, and apply homogamic favoritism among untalented candidates; (b) one organization attracts all talented candidates; and (c) the equilibrium is group-coalition proof[14].

Assume organization 1 starts with a higher quality ($S_2(0) < S_1(0) \leq 2x$)[15] and, say, majoritarian group A, while organization 2 starts with a B majority[16]. If all talented candidates choose organization 1, the dynamics of the state variables are given by:

$$\frac{dS_1}{dt} = \chi(-S_1 + 2x), \qquad \frac{dS_2}{dt} = -\chi S_2, \qquad \frac{dM_1}{dt} = \chi(-M_1 + 1 - x), \qquad \frac{dM_2}{dt} = \chi(-M_2 + 1)$$

Group-coalition proofness for talented B-candidates is satisfied if they would not contemplate a collective deviation to joining organization 2:

$$\int_0^\infty e^{-(r+\chi)t}\Big[[S_1(0) - 2x]\tilde{s}e^{-\chi t} + 2x\tilde{s} + \Big(1 - [M_1(0) - (1-x)]e^{-\chi t} - (1-x)\Big)\tilde{b}\Big]dt$$
$$\geq \int_0^\infty e^{-(r+\chi)t}\Big[[S_2(0) - x]\tilde{s}e^{-\chi t} + x\tilde{s} + [M_2(0) - 1]\tilde{b}e^{-\chi t} + \tilde{b}\Big]dt$$

i.e. if and only if the differential in the initial value proposition exceeds a (turnover-and-interest-rate weighted) long-term loss (benefit if negative):[17]

$$[S_1(0) - S_2(0)]s + [(1 - M_1(0)) - M_2(0)]b \geq \frac{\chi}{r+\chi}[(1-x)b - xs] \tag{6}$$

Pursuing the analysis along these lines, we can show:

Proposition 2. *(Increasing-dominance equilibria) Suppose that organization 1 starts with an A-majority and higher quality ($S_1(0) > S_2(0)$) than organization 2, which starts with a B-majority. There exists $\rho_0 > 0$ such that*

- *for $s/b < \rho_0$, there exists no increasing-dominance equilibrium,*

[13] These notions are in the spirit of Bernheim et al (1987).

[14] Online Appendix E investigates the population-coalition proofness of these equilibria, showing in particular that this property is self-reinforcing over time.

[15] Our analysis applies to any initial qualities $S_j(0) \in [0,1]$, yet for the sake of exposition we assume $S_j(0) \leq 2x$.

[16] Alternatively, insights are unaltered if it starts with an A majority (see Online Appendix E).

[17] The group-coalition proofness condition for talented A-group candidates is a fortiori satisfied when (6) is:

$$\big[S_1(0) - S_2(0)\big]s + \big[M_1(0) - \big(1 - M_2(0)\big)\big]b \geq -\frac{\chi}{r+\chi}\big[(1-2x)b + xs\big].$$

- *for $s/b \geq \rho_0$, there exists an increasing-dominance equilibrium in which all talented candidates join organization 1. There is no other such equilibrium if the initial quality differential is large ($S_1(0) - S_2(0) > x\chi/(r + \chi)$), while for a smaller initial differential and if s/b is greater than some threshold $\rho_1 \geq \rho_0$, there exists another such equilibrium, in which all talented candidates join organization 2.*

5 Conclusion

The paper's two main insights, organizational spirals and voluntary affirmative action, were covered in the introduction. Obvious areas for future research include the introduction of concerns for control (see footnote 10 for an example) as well as a full treatment of competition for talent. On the latter front, it would be interesting to understand when affirmative action can become a norm; if some organizations (say public ones) put more emphasis on diversity, other organizations may need to match that diversity to woo minorities (strategic complementarity) or to the contrary may give up on it, finding it even harder to attract minority talent (strategic substitutability).

The paper focused on diversity and quality as determinants of attractiveness. Another dimension is financial compensation. Lagging departments or firms sometimes throw money at stars to jumpstart a virtuous spiral. But this strategy's success is not a foregone conclusion. Like in this paper, adverse expectations may thwart the attempt. A credible long purse allowing management to do "whatever it takes" may help, as does the choice of a credible leadership. But in the end, quality and diversity will always be determinants of which dynamic the organization will enter.

References

Bernheim, B., Peleg, B., and Whinston, M. D. (1987). "Coalition-Proof Nash Equilibria I. Concepts". *Journal of Economic Theory*, 42(1):1–12.

Board, S., Meyer-ter-Vehn, M., and Sadzik, T. (2019). "Recruiting Talent". R&R at *American Economic Review.*

Cai, H., Feng, H., and Weng, X. (2018). "A Theory of Organizational Dynamics: Internal Politics and Efficiency". *American Economic Journal: Microeconomics*, 10(4):94–130.

Moisson, P.-H. and Tirole, J. (2020). "Cooptation: Meritocracy vs. Homophily in Organizations". mimeo.

Online Appendix

A A preliminary result

A simple result used repeatedly is the following: let $X(t)$ follow $dX/dt = \chi(-X + X^*)$ (with X^* the steady state value). Then $X(t) = (X(0) - X^*)e^{-\chi t} + X^*$, and the PDV of the flow $X(t)dt$ (weighted by time preference and exit probability) is a convex combination of the initial value and the steady state value:

$$\int_0^\infty e^{-(r+\chi)t} X(t)dt = \frac{1}{r+\chi}\Big(\frac{(r+\chi)X(0) + \chi X^*}{r+2\chi}\Big).$$

B Only talented minority candidates have an outside option and $u \leq b/2$

(A) Suppose first that only talented minority candidates have an opportunity cost for joining the organization. [In case (B), both majority and minority talented candidates will have the same opportunity cost for joining the organization (see Online Appendix C).]

Within case (A), we distinguish two subcases depending on the parameters' values:

(A.1) $u \leq b/2$,

(A.2) $u > b/2$, with (A.2.a): $u - b/2 > xs$, and (A.2.b): $u - b/2 < xs$.

(A.1) Suppose $u \leq b/2$. The majority's program writes as

$$\max_{\sigma_0,\sigma_1,\sigma_2} \int_0^{+\infty} e^{-(r+\chi)t} \Big[S_t \tilde{s} + M_t \tilde{b}\Big] dt$$

subject to

(i) if $S_t s - M_t b \geq u - b$,

$$\frac{dM_t}{dt} = \chi\big[-M_t + x(\sigma_1 + 1 - \sigma_2) + (1-2x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS_t}{dt} = \chi\big[-S_t + x(\sigma_1 + \sigma_2)\big]$$

(ii) if $S_t s - M_t b < u - b$,

$$\frac{dM_t}{dt} = \chi\big[-M_t + x\sigma_1 + (1-x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS_t}{dt} = \chi\big[-S_t + x\sigma_1\big]$$

Proposition B.1. ***(Only talented minority candidates have an outside option)*** *Assume (4) is satisfied, and $u \leq b/2$. The following is a solution to the majority's optimal control problem:*

- *(Region 1) If $S_t s - M_t b > u - b$, the majority selects $\sigma_1 = \sigma_2 = \sigma_0 = 1$.*
- *(Region 2) If $S_t s - M_t b = u - b$, the majority selects $\sigma_1 = \sigma_2 = 1$ and $\sigma_0 = \sigma_0^*$.*
- *(Region 3) If $S_t s - M_t b < u - b$ and (M_0, S_0) satisfies (5), the majority selects $\sigma_1 = 1$ and $\sigma_0 = 0$.*
- *(Region 4) If $S_t s - M_t b < u - b$ and (M_0, S_0) does not satisfy (5), the majority selects $\sigma_1 = \sigma_0 = 1$.*

If (4) is not satisfied, then Region 3 is empty, and whenever $S_t s - M_t b < u - b$, the majority selects $\sigma_1 = \sigma_0 = 1$.

Proof. Consider a solution $(\sigma_0, \sigma_1, \sigma_2)$ to the majority's optimal control problem.

That region 2 is absorbing for $\tilde{u} \in [2x\tilde{s} + x\tilde{b}, 2x\tilde{s} + \tilde{b}/2]$ derives from the above discussion. Moreover, for constant controls σ_1 and σ_2, the dynamics of $\tilde{S}_t$ over the sets $\{(u_t, \tilde{S}_t) | u_t + \tilde{u} \leq 2\tilde{S}_t + \tilde{b}\}$ and $\{(u_t, \tilde{S}_t) | u_t + \tilde{u} > 2\tilde{S}_t + \tilde{b}\}$ do not depend on the majority's size M nor on the control σ_0.

Consider region 1, i.e. the set $\{(u_t, \tilde{S}_t) | u_t + \tilde{u} < 2\tilde{S}_t + \tilde{b}\}$. We first note that *region 2 is reached in a finite time from region 1.* Indeed, if region 2 is never reached, our initial assumptions on $\tilde{u}$ imply that $\sigma_0 < 1$ or $\sigma_2 < 1$ (or both) on a non-empty interval, and thus that the majority could strictly improve its welfare by slightly increasing σ_0 (since $\tilde{b} > 0$) or σ_2 (since $\tilde{s} \geq \tilde{b}$) on this interval, still without ever reaching region 2.

Lemma B.2. *For a time $T < \infty$ of arrival in region 2, let $V((M(T), S(T)))$ denote the continuation value function for the majority. Then,*

$$\frac{\partial V}{\partial M(T)}(M(T), S(T)) = b, \qquad \textit{and} \qquad \frac{\partial V}{\partial S(T)}(M(T), S(T)) = s \tag{7}$$

Indeed, using the dynamics of M and S over region 2 yields that for all $t \geq T$,

$$M(t) = \left[M(T) - x - (1-2x)\sigma_0^*\right]e^{-\chi(t-T)} + x + (1-2x)\sigma_0^*,$$
$$S(t) = \left[S(T) - 2x\right]e^{-\chi(t-T)} + 2x$$

Consequently,

$$\begin{aligned} V(M(T),S(T)) &= \int_0^\infty e^{-(r+\chi)t}\tilde{b}\Big([M(T)-x-(1-2x)\sigma_0^*]e^{-\chi t}+x+(1-2x)\sigma_0^*\Big)dt \\ &\quad+\int_0^\infty e^{-(r+\chi)t}\tilde{s}\Big([S(T)-2x]e^{-\chi t}+2x\Big)dt \\ &= M(T)b+S(T)s+\frac{\chi}{r+\chi}\Big([x+(1-2x)\sigma_0^*]b+2xs\Big) \end{aligned}$$

And thus by differentiation,

$$\begin{aligned} \frac{\partial V}{\partial M(T)}(M(T),S(T)) &= b, \\ \frac{\partial V}{\partial M(T)}(M(T),S(T)) &= s \end{aligned}$$

The majority's optimal control problem in region 1. The majority solves:

$$\max_{\sigma_0,\sigma_1,\sigma_2,T} \quad \left\{\int_0^T e^{-(r+\chi)t}\Big[\tilde{s}S(t)+\tilde{b}M(t)\Big]dt+e^{-(r+\chi)T}V((M(T),S(T)))\right\}$$

subject to (8) and (9), which are respectively the final time constraint

$$sS(T)-bM(T)=u-b \tag{8}$$

and the state dynamics

$$\frac{dM}{dt}=\chi\big[-M+x(\sigma_1+1-\sigma_2)+(1-2x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS}{dt}=\chi\big[-S+x(\sigma_1+\sigma_2)\big] \tag{9}$$

So the Hamiltonian writes as

$$H\equiv e^{-(r+\chi)t}\big[\tilde{s}S+\tilde{b}M\big]+\chi p(t)\big[-M+x(\sigma_1+1-\sigma_2)+(1-2x)\sigma_0\big]+\chi q(t)\big[-S+x(\sigma_1+\sigma_2)\big]$$

Hence, requiring that

$$-\frac{dp}{dt}=\frac{\partial H}{\partial M}=\tilde{b}e^{-(r+\chi)t}-\chi p, \qquad \text{and} \qquad -\frac{dq}{dt}=\frac{\partial H}{\partial S}=\tilde{s}e^{-(r+\chi)t}-\chi q$$

and, letting $\psi>0$ be the multiplier for the final time constraint (8),

$$p(T)=e^{-(r+\chi)T}\frac{\partial V}{\partial M}(M(T),S(T))-\psi b, \qquad \text{and} \qquad q(T)=e^{-(r+\chi)T}\frac{\partial V}{\partial S}(M(T),S(T))+\psi s$$

which together with (7) imply that

$$p(t) = be^{-(r+\chi)t} - \psi be^{-\chi(T-t)}, \qquad \text{and} \qquad q(t) = se^{-(r+\chi)t} + \psi se^{-\chi(T-t)},$$

the Hamiltonian's partial derivatives write as

$$\begin{cases} \dfrac{\partial H}{\partial \sigma_0} = \chi\Big(e^{-(r+\chi)t} - \psi e^{-\chi(T-t)}\Big)(1-2x)b, \\ \dfrac{\partial H}{\partial \sigma_1} = \chi e^{-(r+\chi)t}x(s+b) + \psi\chi e^{-\chi(T-t)}x(s-b), \\ \dfrac{\partial H}{\partial \sigma_2} = \chi e^{-(r+\chi)t}x(s-b) + \psi\chi e^{-\chi(T-t)}x(s+b) \end{cases} \tag{10}$$

Pontryagin's maximum principle with variable horizon thus yields that the optimal control σ satisfies $\sigma_1 = \sigma_2 = 1$, and the sum of the Hamiltonian and the partial derivative of the final cost with respect to the final time, evaluated at the final time T, must be nil:

$$\begin{aligned} & e^{-(r+\chi)T}\big[\tilde{s}S(T) + \tilde{b}M(T)\big] + \chi p(T)\big[-M(T) + x(\sigma_1 + 1 - \sigma_2) + (1-2x)\sigma_0\big] \\ & \qquad + \chi q(T)\big[-S(T) + x(\sigma_1 + \sigma_2)\big] \\ & = (r+\chi)e^{-(r+\chi)T}V(M(T), S(T)) \end{aligned}$$

i.e. by using the final time constraint (8), replacing the controls σ_1 and σ_2 with their optimal values $\sigma_1 = \sigma_2 = 1$, and rearranging,

$$e^{-(r+\chi)T}(1-2x)(\sigma_0 - \sigma_0^*)b = \psi\big[2(1-x)b + (1-2x)(\sigma_0 - \sigma_0^*)b\big]$$

Hence, $\psi < e^{-(r+\chi)T}$, and thus $\sigma_0 = 1$.[18]

The majority's optimal control problem in regions 3 and 4. We first suppose region 2 is reached in a finite time T and apply the same arguments as above in order to derive the optimal controls and finite time for region 2 to be reached. We then compare this (optimal) value of reaching region 2 in a finite time to the (optimal) value of never reaching it. The cutoff condition – which is condition (5) in the text – draws the line between regions 3 and 4.

(i) Suppose region 2 is reached at time $T < \infty$. Then (7) holds. The majority's optimiza-

[18] An intuition for $\psi < e^{-(r+\chi)T}$ is that the continuation value upon reaching region 2 is lower than the value of being in region 1. Conversely, in the Pontryagin maximization problem in regions 3 and 4, $\psi > e^{-(r+\chi)T}$ as the continuation value upon reaching region 2 is higher than the value of being in region 3.

tion problem writes as

$$\max_{\sigma_0,\sigma_1,\sigma_2,T} \quad \left\{ \int_0^T e^{-(r+\chi)t}\Big[\tilde{s}S(t)+\tilde{b}M(t)\Big]dt + e^{-(r+\chi)T}V((M(T),S(T)))\right\}$$

subject to (11) and (12) which are respectively the final time constraint

$$sS(T)-bM(T)=u-b \tag{11}$$

and the state dynamics

$$\frac{dM}{dt}=\chi\big[-M+x\sigma_1+(1-x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS}{dt}=\chi\big[-S+x\sigma_1\big] \tag{12}$$

The Hamiltonian writes

$$H \equiv e^{-(r+\chi)t}\big[\tilde{s}S+\tilde{b}M\big]+\chi p(t)\big[-M+x\sigma_1+(1-x)\sigma_0\big]+\chi q(t)\big[-S+x\sigma_1\big]$$

Hence, requiring that

$$-\frac{dp}{dt}=\frac{\partial H}{\partial M}=\tilde{b}e^{-(r+\chi)t}-\chi p, \qquad \text{and} \qquad -\frac{dq}{dt}=\frac{\partial H}{\partial S}=\tilde{s}e^{-(r+\chi)t}-\chi q$$

and, letting $\psi>0$ be the multiplier for the final time constraint (11),

$$p(T)=e^{-(r+\chi)T}\frac{\partial V}{\partial M}(M(T),S(T))-\psi b, \qquad \text{and} \qquad q(T)=e^{-(r+\chi)T}\frac{\partial V}{\partial S}(M(T),S(T))+\psi s$$

which together with (7) imply that

$$p(t)=be^{-(r+\chi)t}-\psi be^{-\chi(T-t)}, \qquad \text{and} \qquad q(t)=se^{-(r+\chi)t}+\psi se^{-\chi(T-t)},$$

the Hamiltonian's partial derivatives write as

$$\begin{cases} \dfrac{\partial H}{\partial \sigma_0}=\chi(1-x)\Big(be^{-(r+\chi)t}-\psi be^{-\chi(T-t)}\Big), \\ \dfrac{\partial H}{\partial \sigma_1}=\chi x\Big(e^{-(r+\chi)t}(s+b)+\psi\chi e^{-\chi(T-t)}(s-b)\Big) \end{cases} \tag{13}$$

Pontryagin's maximum principle with variable horizon yields that the optimal control σ satisfies $\sigma_1=1$, and the sum of the Hamiltonian and the partial derivative of the final cost with respect to the final time, evaluated at the final time T, must be nil:

$$\begin{aligned}&e^{-(r+\chi)T}\big[\tilde{s}S(T)+\tilde{b}M(T)\big]+\chi p(T)\big[-M(T)+x\sigma_1+(1-x)\sigma_0\big]+\chi q(T)\big[-S(T)+x\sigma_1\big]\\&\qquad=(r+\chi)e^{-(r+\chi)T}V(M(T),S(T))\end{aligned}$$

i.e. by using the final time constraint (11), replacing the control σ_1 with its optimal value ($\sigma_1 = 1$), and rearranging,

$$e^{-(r+\chi)T}\Big[u-(1-x)(1-\sigma_0)b-3xs\Big]=\psi\Big[u-(1-x)(1-\sigma_0)b-xs\Big]$$

Since we assumed that $u < 3xs$ (which is a necessary condition for region 2 to exist, see condition (3) in the text), the LHS is always negative. Hence, for a solution to exist, it must be that $u < xs + (1-x)b$ (which is condition (4) in the text). And therefore, $\psi > e^{-(r+\chi)T}$, and thus $\sigma_0 = 0$.

(ii) It thus remains to compare the value of reaching region 2 in a finite time with the optimal controls, to the value of never reaching region 2 (which clearly yields $\sigma_1 = \sigma_0 = 1$). This is transcribed in the following condition on the initial state (M_0, S_0) (which is condition (5) in the text):

$$\begin{aligned}&\int_0^T e^{-(r+\chi)t}(1-x)b\big[1-e^{-\chi t}\big]dt+\int_T^\infty e^{-(r+\chi)t}(1-x)b\big[1-e^{-\chi T}\big]e^{-\chi(t-T)}dt\\&\qquad\le\int_T^{+\infty}e^{-(r+\chi)t}\big(3xs-u\big)\big[1-e^{-\chi(t-T)}\big]dt\end{aligned}\tag{14}$$

where $T < \infty$ is the time at which region 2 is reached from an optimal path starting from initial state (M_0, S_0), and is thus given by

$$T \equiv \frac{1}{\chi}\ln\left[\frac{M_0b-S_0s+x(s-b)}{xs+(1-x)b-u}\right]\ge 0$$

Indeed, let $u_t \equiv M_tb + S_ts$ be the majority's flow utility. Starting from a couple (M_0, S_0) such that $S_0s - M_0b < u - b$, the majority's flow utility without affirmative action ($\sigma_0 = 1$) writes as

$$\forall t\ge 0,\qquad u_t^{(4)}=[M_0b+S_0s-xs-b]e^{-\chi t}+xs+b,$$

whereas with full affirmative action ($\sigma_0 = 0$), it writes as

$$\forall t\in[0,T],\qquad u_t^{(3)}=[M_0b+S_0s-xs-xb]e^{-\chi t}+xs+xb,$$

where T is the time at which region 2 is reached and is thus given by: $u_T^{(3)} - 2S_T = b - u$, i.e.

$$[M_0 b - S_0 s + x(s-b)]e^{-\chi T} = xs + (1-x)b - u$$

For any $t > T$, the organization remains in region 2, and the majority's utility thus writes as

$$u_t^{(2)} = [u_T^{(3)} - 2xs - xb - (1-2x)\sigma_0^* b]e^{-\chi(t-T)} + 2xs + xb + (1-2x)\sigma_0^* b$$

where $u_T^{(3)} = [M_0 b + S_0 s - xs - xb]e^{-\chi T} + xs + xb$. The majority's sacrifice in region 3 is then worthwhile if and only if

$$\int_0^\infty e^{-(r+\chi)t} u_t^{(4)} dt \leq \int_0^T e^{-(r+\chi)t} u_t^{(3)} dt + \int_T^\infty e^{-(r+\chi)t} u_t^{(2)} dt$$

Condition (5) obtains by rearranging and using the definition of σ_0^*. □

C General exposition

We assume in the following that condition (4) is satisfied. We describe the dynamics in cases (A.2) and (B). Proofs are delayed to Online Appendix D.

(A.2) Suppose $u > b/2$. The analysis in region 1 is left unchanged. By contrast, region 2 now cuts the vertical axis before the horizontal one : namely, the point $(1/2, \frac{u-b/2}{s})$ is the intersection of region 2 with the vertical axis. The above analysis for regions 3 and 4 is thus altered as some trajectories with $\sigma_1 = 1 - \sigma_0 = 1$ ("full affirmative action") which were previously in region 3, now reach the vertical axis before reaching region 2 [19]. The analysis now depends on the sign of $u - b/2 - xs$.

(A.2.a) If $u - b/2 > xs$, then any "affirmative action" trajectory ($\sigma_0 < 1$) coming from below region 2 and reaching the vertical axis below region 2[20], subsequently converges towards a fixed point ($(1/2, x)$) which is on the vertical axis, yet strictly below region 2. Hence region 2 is never reached, and thus optimality requires that, starting from any point on this trajectory, the majority selects $\sigma_1 = \sigma_0 = 1$. In other words, any such point belongs to region 4.

Moreover, for $u - b/2 > xs$, region 2 is reached in a finite time from an initial state (M_0, S_0) if and only if the full-affirmative action trajectory starting from (M_0, S_0) reaches region 2 in a finite time. In addition, the previous analysis still applies yielding that among the values of

[19] Indeed, any such trajectory aims for $M = x < 1/2$ and $S = x$.

[20] And thus a fortiori any trajectory with a lower degree of affirmative action yet still reaching the vertical axis in a finite time.

σ_0 such that region 2 is reached in a finite time, the lowest one is optimal.

As a consequence, starting from the vertical axis, the frontier between regions 3 and 4 is now given first by the "full-affirmative-action" trajectory ($\sigma_1 = 1 - \sigma_0 = 1$) which cuts the vertical axis in $(1/2, \dfrac{u - b/2}{s})$, until this trajectory reaches the line defined by (5), after which the frontier is given as before by the latter, which is an increasing line parallel to region 2 [21]: Region 3 is the set of initial states below region 2 and above this frontier.

(A.2.b) If $u - b/2 < xs$, then if the organization starts from region 3 and reaches the vertical axis before region 2, it subsequently goes up the vertical axis towards the state $(1/2, x)$. Since this state is strictly above region 2, the latter is reached in a finite time. Yet by choosing a lower intensity of affirmative action ($\overline{\sigma}_0 \geq 0$), the organization can reach the vertical axis at its intersection with region 2. We show in Online Appendix D that among all intensities of affirmative action such that region 2 is reached in a finite time, it is optimal for the majority to choose the lowest possible σ_0 such that region 2 is reached before the vertical axis[22]. As a consequence, the organization engages in full affirmative action ($\sigma_0 = 0$) whenever the latter makes the organization reach region 2 before the vertical axis, and otherwise selects $\overline{\sigma}_0 > 0$ defined as the value for which the organization reaches region 2 on the vertical axis, i.e. at the point $(1/2, \dfrac{u - b/2}{s})$.

[21] Indeed, our previous analysis of the optimal control problem still applies to any point on this trajectory, yielding that among all levels of affirmative action, full affirmative action is optimal. Condition (5) then ensures that full affirmative action is optimal with respect to standard favoritism.

[22] An intuition underlying this result is as follows:

- $\sigma_1 = 1$ is optimal for the same reasons as before,
- consider the (closure of the) set of strategies σ_0 such that region 2 is reached before the vertical axis: the previous analysis applies, yielding that the lowest such σ_0 is optimal.
- consider the (closure of the) set of strategies σ_0 such that the vertical axis is reached before region 2. We observe that (i) all these trajectories ultimately reach region 2 at the same point (i.e. $((1/2, \dfrac{u - b/2}{s}))$, and (ii) the dynamics of S_0 withing a region do not depend on the value of the control. Therefore, all these trajectories reach region 2 at the same time. The result thus follows from the observation that picking the highest possible σ_0 within this set grants the highest homophily flow benefits, without any quality losses nor delay in reaching region 2.

Namely, given the initial state (M_0, S_0), $\overline{\sigma}_0$ is given whenever it exists by[23]

$$\begin{cases} [M_0 - x - (1-x)\overline{\sigma}_0]e^{-\chi\overline{T}} + x + (1-x)\overline{\sigma}_0 = \dfrac{1}{2} \\ [S_0 - x]se^{-\chi\overline{T}} + xs = u - \dfrac{b}{2} \end{cases}$$

It remains to compare, whenever it applies, affirmative action with intensity $\overline{\sigma}_0$ to standard favoritism ($\sigma_1 = \sigma_0 = 1$). It it thus optimal for the organization to aim for region 2 starting from an initial state such that full affirmative action would lead to the vertical axis before region 2, if and only if[24]

$$\int_0^{\overline{T}} e^{-(r+\chi)t}(1-x)b[1-\overline{\sigma}_0](1-e^{-\chi t})dt + \int_{\overline{T}}^{\infty} e^{-(r+\chi)t}(1-x)b[1-\overline{\sigma}_0](1-e^{-\chi\overline{T}})e^{-\chi(t-\overline{T})}dt$$
$$\leq \int_{\overline{T}}^{\infty} e^{-(r+\chi)t}(3xs-u)[1-e^{-\chi(t-\overline{T})}]dt \tag{15}$$

Given an initial state S_0 (and thus given $\overline{T}$), condition (15) with equality uniquely defines $\overline{\sigma}_0$ (and thus gives a unique M_0). Hence, since $\overline{\sigma}_0$ increases with S_0 and decreases with M_0, condition (15) with equality defines an upward-sloping curve in the plane (M, S), which we denote by Γ'. Moreover, since the LHS in (15) decreases with $\overline{\sigma}_0$, any point on the left of Γ' satisfies the condition.

Let Γ_{AA} be the full affirmative-action trajectory ($\sigma_1 = 1 - \sigma_0 = 1$) which cuts the vertical axis in $(1/2, \dfrac{u-b/2}{s})$. The frontier between regions 3 and 4 is now given by the set of points in $\{(M,S)\,|\,M \in [1/2,1], S \in [0,1]\}$ below region 2 and either (i) below line Γ_{AA} and above line Γ', or (ii) above line Γ_{AA} and to the left of the line defined by (5).[25]

(B) We now assume that both the majority's and the minority's talented candidates have the same (normalized) opportunity cost for joining the organization u. Then there may exist an additional region where the organization fails to recruit such candidates (which we refer to as "region 5").

The set of states such that the majority's flow utility equals its outside option is given

[23] Note that the "frontier" defined by $\overline{\sigma}_0 = 0$ (i.e. the set of largest initial majority sizes such that the system has a solution given an initial quality) is a decreasing line in the plane (M, S), given by the set of initial states satisfying

$$\frac{b}{s}\frac{M_0 - x}{S_0 - x} = \frac{\dfrac{b}{2} - x}{u - \dfrac{b}{2} - xs}$$

[24] See Online Appendix D for details.

[25] Indeed, our previous analysis of the optimal control problem still applies to any point above this trajectory, yielding that among all levels of affirmative action, full affirmative action is optimal. Condition (5) then ensures that full affirmative action is optimal with respect to standard favoritism.

by the line $\Gamma \equiv \{(M,S)\,|\,Mb + Ss = u\}$. The line Γ is an upper bound on the frontier between regions 4 and 5. Indeed, for any point below this line, $Mb + Ss < u$ and thus, if the organization remains below Γ, the participation constraint of talented majority candidates is not met. Yet it may be that the organization does not remain below Γ (see below), in which case the frontier between regions 4 and 5 lies strictly below Γ.

Moreover, whenever the organization falls in region 5, it is left with a single control which is the fraction of untalented majority candidates. Yet because talented candidates of both sides have the same outside option, sacrificing homophily is strictly suboptimal for the majority. The state dynamics in region 5 are thus given by

$$\frac{dM}{dt} = \chi(-M+1), \qquad \text{and} \qquad \frac{dS}{dt} = -\chi S$$

Hence any trajectory starting from region 5 converges towards the point $(1,0)$: this point may or may not be interior to region 5 as the line Γ has vertical coordinate $(u-b)/s$ for $M = 1$ (see below).

A necessary condition for region 5 to be non-empty is thus $u > b/2$, i.e. that Γ cross the vertical axis strictly above the horizontal axis. [26]

When talented candidates of both groups have an outside option, the majority's optimal control problem when the organization is on the right of region 2 may differ from when only talented minority candidates have such an option. We refer to Online Appendix D for a detailed description of the phase diagram. We only mention here that for $u \leq b/2$, the participation constraint of talented majority candidates is never binding as they are always guaranteed at least $b/2$ upon joining the organization. Hence for $u \leq b/2$, the above analysis remains unchanged (and region 5 is empty).

D Proof of Proposition 1

(A.2.b.) ***Only talented minority candidates have an opportunity cost for joining the organization, and*** $b/2 < u < xs + b/2$**.** We first establish that, starting from an initial state such that a full affirmative action would lead to the vertical axis strictly below its intersection with region 2, if region 2 is reached in a finite time, then the affirmative action trajectory that reaches region 2 at its intersection with the vertical axis ($\sigma_0 = \overline{\sigma}_0$) is optimal. Yet since some trajectories may reach the vertical axis before region 2, there may be

[26] Moreover, the line Γ and the line defining region 2 reach the vertical axis in the same point, namely $\big(1/2, \frac{u-b/2}{s}\big)$. Indeed, talented candidates of both sides have the same outside option and for $M = 1/2$, they enjoy the same flow utility.

a discontinuity in the dynamics of M. We thus show the result by considering two distinct Pontryagin maximization problems and compare their optimal values.

It can be shown (with Pontryagin arguments on well chosen parameter sets) that $\sigma_1 = 1$ is always optimal. We thus focus on the choice of σ_0. Let (as before) $\overline{\sigma}_0$ be the parameter value such that the trajectory with control $\sigma_0 = \overline{\sigma}_0$ reaches the point $(M, S) = (1/2, \dfrac{u - b/2}{s})$. Hence $\overline{\sigma}_0$ is the lowest parameter value for control σ_0 such that the trajectory reaches region 2 before the vertical axis. Namely, given the initial state (M_0, S_0), $\overline{\sigma}_0$ is given whenever it exists by

$$\begin{cases} [M_0 - x - (1-x)\overline{\sigma}_0]e^{-\chi\overline{T}} + x + (1-x)\overline{\sigma}_0 = \dfrac{1}{2} \\ [S_0 - x]se^{-\chi\overline{T}} + xs = u - \dfrac{b}{2} \end{cases}$$

We thus distinguish two sets of admissible values for the control σ_0:

- For $\sigma_0 \in [\overline{\sigma}_0, 1]$, the previous Pontryagin maximization problem yields that $\overline{\sigma}_0$ is optimal. The organization thus reaches region 2 at time $\overline{T}$ at the point $(1/2, \dfrac{u - b/2}{s})$.
- For $\sigma_0 \in [0, \overline{\sigma}_0]$, the problem writes differently as the vertical axis is reached before region 2. Let $(1/2, S)$ be the point on the vertical axis reached by a given trajectory at time T_1. The continuation value from state $(1/2, S(T_1))$ reached at time T_1, denoted by $V^\dagger(T_1, 1/2, S(T_1))$ writes as

$$\int_0^{T_2-T_1} e^{-(r+\chi)t}\frac{\tilde{b}}{2}dt + \int_0^{T_2-T_1} e^{-(r+\chi)t}\Big[(S(T_1) - x)\tilde{s}e^{-\chi t} + x\tilde{s}\Big]dt$$
$$+ \int_{T_2-T_1}^{\infty} e^{-(r+\chi)t}\Big[\Big(\tilde{u} - 2x\tilde{s} - (x + (1-2x)\sigma_0^*)\tilde{b}\Big)e^{-\chi(t-T_2+T_1)} + 2x\tilde{s} + (x + (1-2x)\sigma_0^*)\tilde{b}\Big]dt$$

where T_2 is given by

$$[S_0 - x]se^{-\chi T_2} + xs = u - \frac{b}{2}$$

The majority's optimization problem writes as

$$\max_{\sigma_0 \in [0, \overline{\sigma}_0], T_1} \left\{ \int_0^{T_1} e^{-(r+\chi)t}\Big[\tilde{s}S(t) + \tilde{b}M(t)\Big]dt + e^{-(r+\chi)T_1}V^\dagger(T_1, 1/2, S(T_1))) \right\}$$

subject to the final time constraint $M(T_1) = 1/2$ and the state dynamics

$$\frac{dM}{dt} = \chi\big[-M + x + (1-x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS}{dt} = \chi\big[-S + x\big]$$

The Hamiltonian writes

$$H \equiv e^{-(r+\chi)t}\big[\tilde{s}S + \tilde{b}M\big] + \chi p(t)\big[-M + x + (1-x)\sigma_0\big] + \chi q(t)\big[-S + x\big]$$

Hence, requiring that

$$-\frac{dp}{dt} = \frac{\partial H}{\partial M} = \tilde{b}e^{-(r+\chi)t} - \chi p, \qquad \text{and} \qquad -\frac{dq}{dt} = \frac{\partial H}{\partial S} = \tilde{s}e^{-(r+\chi)t} - \chi q$$

and, letting $\psi > 0$ be the multiplier for the final time constraint,

$$p(T_1) = \psi, \qquad \text{and} \qquad q(T_1) = e^{-(r+\chi)T_1}\frac{\partial V^\dagger}{\partial S}(T_1, 1/2, S) = e^{-(r+\chi)T_1}\Big(1 - e^{-(r+2\chi)(T_2-T_1)}\Big)s$$

which implies that

$$p(t) = be^{-(r+\chi)t} + \psi e^{-\chi(T_1-t)},$$

the Hamiltonian's partial derivative with respect to σ_0 writes as

$$\frac{\partial H}{\partial \sigma_0} = \chi(1-x)\Big[be^{-(r+\chi)t} + \psi e^{-\chi(T_1-t)}\Big] > 0$$

Hence Pontryagin's maximum principle with variable horizon yields that[27] the optimal control σ_0 must be the highest possible, i.e. $\sigma_0 = \overline{\sigma}_0$.

Therefore, if region 2 is reached in a finite time, then optimality requires $\sigma_0 = \overline{\sigma}_0$ (and $\sigma_1 = 1$) as long as region 2 is not reached.

It thus remains to compare the value of reaching region 2 at its intersection with the vertical axis, namely at the point $(\frac{1}{2}, \frac{u - b/2}{s})$ with the value of standard favoritism. The argument for the optimality condition is similar to the one in case A.1. By construction of $\overline{\sigma}_0$ and $\overline{T}$, the condition for the optimality of level-$\overline{\sigma}_0$ affirmative action with respect to standard

[27] Moreover, the sum of the Hamiltonian and the partial derivative of the final cost with respect to the final time, evaluated at the final time T_1, must be nil, and thus:

$$e^{-(r+\chi)T_1}\left[\frac{\tilde{b}}{2} + S(T_1)\tilde{s}\right] + p(T_1)\big[-1/2 + x + (1-x)\sigma_0\big] + q(T_1)\big[-S(T_1) + x\big]$$
$$= e^{-(r+\chi)T_1}\left[(r+\chi)V^\dagger(T_1, 1/2, S(T_1)) - \frac{\partial V^\dagger}{\partial T_1}(T_1, 1/2, S(T_1))\right],$$

which implies that:

$$\psi = \frac{\big[x - S(T_1)\big](1-\chi)s}{\frac{1}{2} - x - (1-x)\sigma_0}\left[e^{-(r+2\chi)T_1} - e^{-(r+2\chi)T_2}\right] > 0$$

favoritism writes as

$$\begin{aligned}
&\int_0^{\overline{T}} e^{-(r+\chi)t}\Big[[S_0 s + M_0 b - xs - (x+(1-x)\overline{\sigma}_0)b]e^{-\chi t} + xs + (x+(1-x)\overline{\sigma}_0)b\Big]dt \\
&\qquad + \int_{\overline{T}}^{\infty} e^{-(r+\chi)t}\Big[[u - 2xs - (x+(1-2x)\sigma_0^*)b]e^{-\chi(t-\overline{T})} + 2xs + (x+(1-2x)\sigma_0^*)b\Big]dt \\
&\geq \int_0^{\infty} e^{-(r+\chi)t}\Big[[S_0 s + M_0 b - xs - b]e^{-\chi t} + xs + b\Big]dt
\end{aligned}$$

which yields (15) after rearranging.

(B) ***Both majority and minority talented candidates have an opportunity cost for joining the organization.*** We first provide an intuition for the results. Consider an "affirmative action" trajectory that reaches the interior of region 5 before the vertical axis. Such a trajectory henceforth converges towards $(1, 0)$, possibly exiting region 5 towards region 4 in a finite time. Hence, because of discounting, this strategy is dominated by "standard favoritism" from $t = 0$ onward, which leads to a weakly more favourable steady state. Moreover, consider an initial state (M_0, S_0) such that the full-affirmative action trajectory ($\sigma_0 = 0$) starting from this state, reaches region 2 in a finite time. Consider any less-than-full affirmative action trajectory ($\sigma_0 > 0$) starting from the same initial state (M_0, S_0). Then,

- if this less-than-full affirmative action trajectory does not reach region 2 in a finite time, it is clearly dominated by "standard favoritism" ($\sigma_0 = 1$ and if possible $\sigma_1 = 1$).
- if this less-than-full affirmative action trajectory reaches region 2 in a finite time, the above analysis applies, yielding that this trajectory is dominated by a full-affirmative action trajectory if it reaches region 2 before the vertical axis, or by the affirmative action trajectory such that region 2 is reached at its intersection with the vertical axis.

Hence the initial state (M_0, S_0) belongs to region 3 only if either (5) or (15) hold, and belongs to regions 4 or 5 otherwise.

As with case (A), we distinguish within case (B) two subcases:

(B.1) $u \leq b/2$,

(B.2) $u > b/2$, with (B.2.a): $u - b/2 \geq xs$, and (B.2.b): $u - b/2 < xs$.

(B.1) Suppose $u \leq b/2$. Then region 5 is empty. The above analysis of case (A.1) is unchanged: the participation constraint of talented majority candidates never binds as they are always guaranteed at least $b/2$ upon joining the organization.

(B.2) Suppose $u > b/2$.

Whenever $u \leq b/2$, the frontier between regions 5 and 4 is given by the set of states (violating (15) if $u - b/2 < xs$) such that

$$\int_0^\infty e^{-(r+\chi)t}\Big[[S_0 - x]\tilde{s}e^{-\chi t} + x\tilde{s} + [M_0 - 1]\tilde{b}e^{-\chi t} + \tilde{b}\Big]dt = \int_0^\infty e^{-(r+\chi)t}\tilde{u}dt \qquad (16)$$

Since the LHS in (16) is strictly increasing with respect to S_0 and M_0, the frontier between regions 5 and 4 has a decreasing slope in the plane (M, S). As a consequence, the state $(M, S) = (1, 0)$ is interior to region 5 if and only if

$$\int_0^\infty e^{-(r+\chi)t}\Big[x\tilde{s}(1 - e^{-\chi t}) + \tilde{b}\Big]dt < \int_0^\infty e^{-(r+\chi)t}\tilde{u}dt,$$

i.e. if

$$u > b + \frac{\chi}{r + 2\chi}xs \qquad (17)$$

Hence, if (17) holds, then whenever the organization starts in region 5, it converges to the steady state $(M, S) = (1, 0)$. There is no escape from region 5.

By contrast, if (17) does not hold, then the point $(1, 0)$ is outside region 5. (Put differently, the frontier between regions 4 and 5 crosses the horizontal axis before reaching $M = 1$). Hence any trajectory from region 5 exits the region, and reaches either region 3 or region 4 in a finite time. If it reaches the latter, it then converges towards region 4's steady-state $(1, x)$.[28]

(B.2.a) If $u - b/2 \geq xs$, then region 5 and region 3 have no shared boundary[29]. Region 5 is given by the set of states below its boundary with region 4 (see case B.2.c below). As a consequence, region 5 is non-empty if and only if the initial state $(1/2, 0)$ satisfies (see (16) below)

$$\int_0^\infty e^{-(r+\chi)t}\Big[x\tilde{s}(1 - e^{-\chi t}) - \frac{\tilde{b}}{2}e^{-\chi t} + \tilde{b}\Big]dt < \int_0^\infty e^{-(r+\chi)t}\tilde{u}dt,$$

i.e. if and only if

$$\chi xs + (r + 3\chi)\frac{b}{2} < (r + 2\chi)u$$

[28] The condition $u < xs + (1 - x)b$ (condition (10) in the paper) implies that the fixed point of region 4 is interior to the region ($xs + b > u$).

[29] Indeed, region 5 lies below the line Γ which is decreasing, while region 3 lies above the full-affirmative-action trajectory going reaching region 2 on the vertical axis, which is increasing. [Recall that the line Γ crosses the vertical axis in $\big(1/2, \frac{u - b/2}{s}\big)$.]

In particular, region 5 is thus non-empty for any χ sufficiently low. If in addition $xs + 3b/2 > 2u$, it is also non-empty for any χ sufficiently high.

(B.2.b) If $u - b/2 < xs$, then regions 5 and 3 may have a shared boundary. Region 5 lies below the curve Γ, while for any initial state (M_0, S_0) below Γ, region 3 is defined by (15). Hence the boundary between region 5 and region 3 is given by the set of initial states (M_0, S_0) (satisfiying (15) with equality) such that

$$\begin{aligned}
&\int_0^{\overline{T}} e^{-(r+\chi)t}\Big[[S_0 - x]\tilde{s}e^{-\chi t} + x\tilde{s} + [M_0 - x - (1-x)\overline{\sigma}_0]\tilde{b}e^{-\chi t} + x\tilde{b} + (1-x)\overline{\sigma}_0\tilde{b}\Big]dt \\
&\quad + \int_{\overline{T}}^{\infty} e^{-(r+\chi)t}\Big[\Big(\tilde{u} - \frac{\tilde{b}}{2} - 2x\tilde{s}\Big)e^{-\chi(t-\overline{T})} + 2x\tilde{s} \\
&\qquad\qquad + \Big(\frac{\tilde{b}}{2} - (x + (1-2x)\sigma_0^*)\tilde{b}\Big)e^{-\chi(t-\overline{T})} + [x + (1-2x)\sigma_0^*]\tilde{b}\Big]dt \\
&= \int_0^{\infty} e^{-(r+\chi)t}\tilde{u}dt
\end{aligned} \tag{18}$$

where $\overline{T} > 0$, $\overline{\sigma}_0 \in [0,1]$ are given whenever they exist[30] by

$$\begin{cases}
[M_0 - x - (1-x)\overline{\sigma}_0]e^{-\chi\overline{T}} + x + (1-x)\overline{\sigma}_0 = \dfrac{1}{2} \\[2ex]
[S_0 - x]se^{-\chi\overline{T}} + xs = u - \dfrac{b}{2}
\end{cases}$$

The LHS in (18) strictly increases with respect to M_0, and for $\overline{T} \ll 1$ (i.e. $S_0 s$ close to $u - b/2$), as well as for $\overline{T} \gg 1$ (i.e. S_0 close to 0 and $u - b/2$ close to xs), with respect to S_0.[31]. Therefore, the frontier between regions 3 and 5 has a decreasing slope in the plane (M, S)

[30] Recall that the "frontier" defined by $\overline{\sigma}_0 = 0$ (i.e. the set of largest initial majority size such that the system has a solution given an initial quality) is a decreasing line in the plane (M, S), given by the set of initial states satisfying

$$\frac{b}{s}\frac{M_0 - x}{S_0 - x} = \frac{\dfrac{b}{2} - x}{u - \dfrac{b}{2} - xs}$$

[31] Indeed, explicit computations yield

$$\begin{aligned}
\frac{\partial LHS}{\partial M_0} &= b\big[1 - e^{-(r+2\chi)\overline{T}}\big] - \frac{\tilde{b}e^{-\chi\overline{T}}}{1 - e^{-\chi\overline{T}}}\left(\frac{1}{r+\chi}\big[1 - e^{-(r+\chi)\overline{T}}\big] + \frac{1}{r+2\chi}\big[1 - e^{-(r+2\chi)\overline{T}}\big]\right) \\
&= \frac{b}{(r+\chi)[1 - e^{-\chi\overline{T}}]}\Big[(r+\chi) - (r+2\chi)e^{-\chi\overline{T}} + \chi e^{-(r+2\chi)\overline{T}}\Big] > 0
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{\partial LHS}{\partial S_0} = s\big[1 - e^{-(r+2\chi)\overline{T}}\big] + \frac{1}{r+\chi}\frac{1}{\big[1 - e^{-\chi\overline{T}}\big]^2}\frac{1}{S_0 - x}\Big[&(r+\chi)\Big(2u - 4xs - b\Big)\big[1 - e^{-\chi\overline{T}}\big]^2 e^{-(r+\chi)\overline{T}} \\
&+ \Big(M_0 - \frac{b}{2}\Big)e^{-\chi\overline{T}}\big[\chi - (r+2\chi)e^{-(r+\chi)\overline{T}} + (r+\chi)e^{-(r+2\chi)\overline{T}}\big]\Big]
\end{aligned}$$

whenever (i) $S_0 s$ is close to $u - b/2$, or (ii) S_0 is close to 0 (with $u - b/2$ close to xs).

As a consequence, if the state $(M_0, S_0) = (1/2, 0)$ satisfies (15)[32], then region 5 is non-empty if it includes the state $(1/2, 0)$, i.e. if

$$\begin{aligned}
&\int_0^{\overline{T}_0} e^{-(r+\chi)t}\Big[x\tilde{s}[1-e^{-\chi t}] + \frac{\tilde{b}}{2}\Big]dt \\
&\quad + \int_{\overline{T}_0}^{\infty} e^{-(r+\chi)t}\bigg[\Big(\tilde{u} - \frac{\tilde{b}}{2} - 2x\tilde{s}\Big)e^{-\chi(t-\overline{T}_0)} + 2x\tilde{s} \\
&\qquad\qquad + \Big(\frac{\tilde{b}}{2} - (x + (1-2x)\sigma_0^*)\tilde{b}\Big)e^{-\chi(t-\overline{T}_0)} + [x + (1-2x)\sigma_0^*]\tilde{b}\bigg]dt \\
&< \int_0^{\infty} e^{-(r+\chi)t}\tilde{u}dt
\end{aligned}$$

where $\overline{T}_0$ is given by

$$xs[1 - e^{-\chi\overline{T}_0}] = u - \frac{b}{2}, \qquad \text{i.e.} \qquad \overline{T}_0 = \frac{1}{\chi}\ln\left(\frac{xs}{xs - u - b/2}\right)$$

The above condition writes after rearranging (assuming $xs > 0$):

$$(r+\chi) - (r+2\chi)e^{-\chi\overline{T}_0} - 2\chi e^{-(r+\chi)\overline{T}_0} + (2r+\chi)e^{-(r+2\chi)\overline{T}_0} > 0$$

Hence in particular, region 5 is non-empty for χ sufficiently close to 0 and for χ sufficiently high, i.e. if turnover is sufficiently low or sufficiently high. The intuition underlying this result is that when turnover is too low, the organization fails to renew its composition fast enough,

i.e. after rearranging,

$$\begin{aligned}
(r+\chi)\big[1-e^{-\chi\overline{T}}\big]^2(S_0 - x)\frac{\partial LHS}{\partial S_0} &= \big[1-e^{-\chi\overline{T}}\big]^2(r+\chi)\left[\Big(u - \frac{b}{2} - xs\Big)e^{\chi\overline{T}} + (u-3xs)e^{-(r+\chi)\overline{T}} - \frac{b}{2}e^{-(r+\chi)\overline{T}}\right] \\
&\quad + \Big(M_0 b - \frac{b}{2}\Big)e^{-\chi\overline{T}}\big[\chi - (r+2\chi)e^{-(r+\chi)\overline{T}} + (r+\chi)e^{-(r+2\chi)\overline{T}}\big]
\end{aligned}$$

Therefore, since $u - b/2 < xs$, $u < 3xs$, $S_0 < x$, and $M_0 < e^{\chi\overline{T}}\big[1/2 - x\big] + x$, we have that $\dfrac{\partial LHS}{\partial S_0} > 0$ for $\overline{T} \ll 1$ (using a second-order Taylor expansion), as well as for $\overline{T} \gg 1$.

[32] $(1/2, 0)$ satisfies (15) if and only if

$$\int_0^{\overline{T}_0} e^{-(r+\chi)t}\frac{b}{2}(1-e^{-\chi t})dt + \int_{\overline{T}_0}^{\infty} e^{-(r+\chi)t}\frac{b}{2}\big(1-e^{-\chi\overline{T}_0}\big)e^{-\chi(t-\overline{T}_0)}dt \le \int_{\overline{T}_0}^{\infty} e^{-(r+\chi)t}(3xs-u)\big[1-e^{-\chi(t-\overline{T}_0)}\big]dt,$$

i.e. if and only if

$$\Big(3xs + \frac{b}{2} - u\Big)e^{-(r+\chi)\overline{T}_0} \ge \frac{b}{2}$$

where $\overline{T}_0$ is given by

$$xs\big[1-e^{-\chi\overline{T}_0}\big] = u - \frac{b}{2}$$

whereas when turnover is too high, members are likely to quit the organization before they could reap the benefits of membership.

By contrast, if the state $(M_0, S_0) = (1/2, 0)$ violates (15), then a necessary and sufficient condition for region 5 to be non-empty is given by the condition stated in case (B.2.a), namely

$$\chi xs + (r + 3\chi)\frac{b}{2} < (r + 2\chi)u$$

Again, region 5 is non-empty for any χ sufficiently low. If in addition $xs + 3b/2 > 2u$, it is also non-empty for any χ sufficiently high.

E Proof of Proposition 2

The dynamics of quality dominance. Inequality (6) is the non-profitability condition for a collective deviation by all talented B-candidates from joining organization 1 to joining organization 2. If (6) is satisfied at date 0, then $[S_1 - S_2]$ converges towards $2x$, while $[M_2 - (1 - M_1)]$ converges towards $(1 - x)$. Hence if (6) is satisfied at time 0, then by convexity, it is satisfied at any later time $t > 0$ if and only if the steady state satisfies (6), i.e. if and only if[33]

$$2xs - (1 - x)b \geq \frac{\chi}{r + \chi}\big[(1 - x)b - xs\big], \tag{19}$$

which is equivalent to $(2r + 3\chi)xs \geq (r + 2\chi)(1 - x)b$.

If (19) is violated, then there is no quality dominance in the long run, and the quality and majority sizes of both organizations follow the same dynamics and thus converge towards the same values (resp. x and 1) as talented candidates split between the two organizations (A-group ones joining organization 1, and B-group ones joining organization 2).[34]

By contrast, if (6) and (19) hold, then whenever the initial state verifies (6), $[S_1 - S_2]$ converges towards $2x$, while $[M_2 - M_1]$ converges towards x: there is quality dominance in the long run, i.e. one organization converges to a diverse, high-quality organization, while the other ends up being fully homogenous and without any talent.

Let $\Delta U \equiv \big[(S_1 - S_2)s + (1 - M_1 - M_2)b\big]$ be the difference in the utility of talented B-group candidates from joining organization 1 instead of organization 2. We refer to ΔU as the

[33] Talented A-group candidates always prefer joining organzation 1 if they do so at time 0 as the steady state satisfies:

$$2xs + (1 - x)b \geq -\frac{\chi}{r + \chi}\big[(1 - 2x)b + xs\big]$$

[34] Indeed, if (19) is violated, then there exists a later (finite) time at which (6) is violated: talented B-group candidates now choose organization 2 from that date onwards. Hence, because decisions are anticipated, talented B-group candidates should start joining organization 2 strictly before that date. By induction, talented B-group candidates should thus join organization 2 starting from date 0.

"comparative advantage" of organization 1 with respect to organization 2 from the perspective of (talented) B-group candidates. Condition (6) can thus be written as

$$\Delta U(0) \geq \frac{\chi}{r+\chi}[(1-x)b - xs]$$

If (6) and (19) hold, then the dynamics of ΔU are given by

$$\frac{d}{dt}\Delta U = \chi\big[-\Delta U + 2xs - (1-x)b\big]$$

Hence the comparative advantage of organization 1 increases over time if and only if $\Delta U(0) \leq 2xs - (1-x)b$. (Note that (19) implies that $2xs \geq (1-x)b$.)

Group-coalition proofness: Applying the group-deviation criterion, a necessary condition for organization 1 to be increasingly dominant is that all talented B-candidates prefer joining organization 1 to collectively deviating to organisation 2; and symmetrically for organisation 2, for which A-candidates would be most eager to deviate (the "weakest link"). This gives us two necessary conditions for the co-existence of two increasing-dominance equilibria[35]

$$\begin{cases} [S_1(0) - S_2(0)]s + [1 - M_1(0) - M_2(0)]b \geq \dfrac{\chi}{r+\chi}[(1-x)b - xs] \\ [S_2(0) - S_1(0)]s + [1 - M_2(0) - M_1(0)]b \geq \dfrac{\chi}{r+\chi}[(1-x)b - xs] \end{cases}$$

i.e. if and only if

$$[S_2(0) - S_1(0)]s + [1 - M_1(0) - M_2(0)]b \geq \frac{\chi}{r+\chi}[(1-x)b - xs]$$

Hence, let ρ_0 be given by

$$\rho_0 \equiv \max\left\{\frac{r+2\chi}{2r+3\chi}\frac{1-x}{x};\ \left(\frac{\chi}{r+\chi}(1-x) + M_1(0) + M_2(0) - 1\right)\Big/\left(\frac{\chi}{r+\chi}x + S_1(0) - S_2(0)\right)\right\},$$

and, if $[x\chi/(r+\chi) + S_2(0) - S_1(0)] > 0$, define ρ_1 as

$$\rho_1 \equiv \max\left\{\rho_0;\ \left(\frac{\chi}{r+\chi}(1-x) + M_1(0) + M_2(0) - 1\right)\Big/\left(\frac{\chi}{r+\chi}x - S_1(0) + S_2(0)\right)\right\},$$

The following existence regions obtain, depending on the value of s/b,

[35] As noted in the text, taking as given that talented B-group (resp. A-group) candidates choose organization 1 (resp. 2), talented A-group (resp. B-group) best-reply by choosing the same organization, i.e. organization 1 (resp.2) – that is, the organization where they are the majority. Hence the condition for talented candidates of a given group to join the organization where they are not the majority is necessary and sufficient for the existence of an increasing-dominance equilibria. The first (resp. second) equation thus gives the non-profitability condition for a deviation by talented B-candidates (resp. A-candidates) towards joining organization 2 (resp. 1) when talented candidates from the other group join organization 1 (resp. 2).

- for $s/b < \rho_0$, there exists no increasing-dominance equilibrium,
- if $[x\chi/(r+\chi) + S_2(0) - S_1(0)] > 0$, then for $\rho_0 \leq s/b < \rho_1$, there exists a single increasing-dominance equilibrium (which is the one in which all talented candidates join organization 1) – note that this range may be empty –, while for $\rho_1 \leq s/b$, there exist two increasing-dominance equilibria.
- if $[x\chi/(r+\chi) + S_2(0) - S_1(0)] \leq 0$, then for $s/b \geq \rho_0$, there exists a single increasing-dominance equilibrium (which is the one in which all talented candidates join organization 1).

Remark: Alternative assumption on initial majorities. If organization 2 starts with an A-majority, then the equilibrium in which all talented candidates join organization 1 exists if and only if[36]

$$\begin{cases} [S_1(0) - S_2(0)]s + [M_2(0) - M_1(0)]b \geq -\dfrac{\chi}{r+\chi}xs \\ [S_1(0) - S_2(0)]s + [M_1(0) - M_2(0)]b \geq -\dfrac{\chi}{r+\chi}x(s-b) \end{cases}$$

Similarly, the equilibrium in which all talented candidates join organization 2 exists if and only if the above system holds when switching the indices 1 and 2. Hence the two increasing-dominance equilibra coexist if and only if their initial states are sufficiently close.

Population-coalition proofness of the increasing-dominance equilibria. By construction, the above equilibria are immune to a joint deviation by talented candidates of a given group. The equilibrium in which all talented candidates join organization 1 is always immune to a deviation by all talented candidates[37], whereas the equilibrium in which all talented candidates join organization 2 is immune to a deviation by all talented candidates if and only if talented B-candidates would not support an overall deviation to organisation 1 (they are the weakest

[36] Note that the steady state satisfies the above conditions as for any $s/b \geq 1$,

$$2xs + xb \geq -\frac{\chi}{r+\chi}xs, \qquad \text{and} \qquad 2xs - xb \geq -\frac{\chi}{r+\chi}x(s-b)$$

[37] The deviation by all talented candidates is strictly profitable for talented candidates from both groups if and only if

$$\begin{cases} \big[S_1(0) - S_2(0)\big]s + \big[M_1(0) + M_2(0) - 1\big]b < -\dfrac{\chi}{r+\chi}(1-2x)b \\ \big[S_1(0) - S_2(0)\big]s - b\big[M_1(0) + M_2(0) - 1\big]b < \dfrac{\chi}{r+\chi}(1-2x)b \end{cases}$$

link for such a deviation)[38]

$$[S_2(0) - S_1(0)]s + [M_1(0) + M_2(0) - 1]b \geq -\frac{\chi}{r+\chi}(1-2x)b \tag{20}$$

Therefore, the equilibrium in which all talented candidates join organization 1 is population-coalition proof whenever it exists (and remains so at any later date), while by contrast, the equilibrium in which all talented candidates join organization 2 is population-coalition proof whenever it exists if and only if (20) is satisfied. In other words, this equilibrium is population-coalition proof if and only if the initial additional homophily benefit for talented B-group candidates (at least) compensates the initial quality loss in choosing organization 2 instead of organization 1. Moreover, since in the equilibrium in which all talented candidates join organization 2, the LHS in (20) converges to $2xs + (1-x)b > 0$ [39], this equilibrium remains population-coalition proof if it is so at date 0, and becomes population-coalition proof past a finite time (and remains so henceforth) if it is not already at time 0.

[38] The deviation by all talented candidates is strictly profitable for talented candidates from both groups if and only if

$$\begin{cases} \big[S_2(0) - S_1(0)\big]s - \big[M_1(0) + M_2(0) - 1\big]b < \dfrac{\chi}{r+\chi}(1-2x)b \\ \big[S_2(0) - S_1(0)\big]s + \big[M_1(0) + M_2(0) - 1\big]b < -\dfrac{\chi}{r+\chi}(1-2x)b \end{cases}$$

In particular, since we assumed $S_1(0) > S_2(0)$, talented A-group candidates always strictly benefit from such a collective deviation. Hence the equilibrium in which all talented candidates join organization 2 is immune to a deviation by all talented candidates if and only if the latter is unprofitable for talented B-group candidates.

[39] The observation that in that equilibrium, $(S_2 - S_1)$ converges to $2xs \geq 0$ would also yield the result.