Cooptation:
Meritocracy vs. Homophily in Organizations

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

November 19, 2019

Abstract

The paper analyzes the Markovian dynamics, the entrenchment (hiring discrimination and glass ceiling) and the welfare properties of an organization whose members' cooptation decisions are driven by two motives: quality and homophily. Its running theme is that meritocracy is fragile. Yet, to avoid depriving itself of its talent pool an organization may voluntarily engage in affirmative action. Finally, the paper investigates which policy interventions among mandated affirmative action, quality assessment exercises, and the overruling of majority decisions may have unintended consequences for the minority whom they are meant to benefit.

Keywords: Cooptation, organizations, Markov games, meritocracy, glass ceiling, virtuous and vicious spirals, affirmative action, assessment exercises.

JEL Codes: D7, C73, D02, M5.

[†]The authors are grateful to Daron Acemoglu, Guido Friebel, Bob Gibbons, Johannes Hörner, Alessandro Pavan, Martin Peitz, Jérôme Renault, Patrick Rey and to participants at presentations at the Bonn-Mannheim CRC conference (Mainz), Stanford University (Accounting Research Conference), the New Economic School (Moscow), the Berlin School of Economics, the MIT Organizational Economics seminar and the Toulouse School of Economics for helpful suggestions. 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 selection of new members of a board of directors, a corporation, a cooperative, a trade or monetary union, an academic department or a polity, underlies institutional dynamics and determines whether the organization succeeds or is consigned to oblivion. New members most often are coopted[1], occasionally under constraints imposed by internal rules or external intervention.

In coopting new members, existing members are safely predicted to pursue their own agenda, raising the question of whether the organization takes on a life of its own or fulfills its primary mission. The process of cooptation indeed gives rise to three types of externalities: onto minority members, whose voice may not be heard; onto potential members, who may not benefit from equal opportunity and meritocracy; and onto society/third parties whom the organization is meant to serve. Will the organization go astray by reveling in clubishness and contravening meritocracy? If so, can we think of interventions that serve better either its members or its mission? Can the organization by itself engage in costly policies that will later avert a slippery slope? These are examples of questions that this paper is meant to address.

The paper accordingly analyzes the Markovian dynamics, the discrimination in hiring and promotion, and the welfare properties of an organization whose members' cooptation decisions are driven by two motives: quality and homophily.

Our running theme is that meritocracy is fragile. A preference for in-group membership (along gender, religion, ethnicity, politics, scientific field or approach, values, friendship, class loyalty or another dimension) creates a benefit from control and leads to various degrees of violations of meritocracy. Strong forms of entrenchment are to be expected when (a) members are better at assessing the ability of in-group candidates than that of out-group ones, or (b) absenteeism or an imperfect identification of group allegiance can generate unexpected switches in the majority, or else (c) homophily benefits are paramount.

A small increase in homophily benefits or in patience may lead the majority to opt for entrenchment. Outside options add to the fragility: small variations in the initial quality or diversity of the organization (due to staffing disruptions, technological shocks, globalization...) may lead to virtuous or vicious spirals and totally 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 will not attract talented minorities, and possibly even stop at some point 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.

We then turn to hierarchical organizations and to the possibility that women (or minorities) experience difficulties in rising beyond a certain level in the hierarchy. Even if male dominance and favoritism contribute to discrimination against women, it is not a priori obvious that they imply a lower rate of

[1] We focus on "cooptation" in the sense of "periodic selection of new members to join the group". A second and equally important acception of "cooptation", associated with Selznick (1948, 1949), argues that absorbing new elements in an organization can be a means of averting threats to its stability or existence. We refer to the literature building on Acemoglu and Robinson (2000)'s celebrated analysis on the extension of the franchise to avoid upheaval (threat-averting cooptation involves the entire threatening group in Acemoglu-Robinson, and only a sub-group in Bertocchi-Spagat 2001). We briefly discuss the link between the two meanings of "cooptation" in the conclusion.

promotion for women and therefore a glass ceiling. We show nonetheless that a glass ceiling results from control being located at the senior level. This operates through two channels: 1) Concern for control: control allows the dominant group to engage in favoritism. Because it is located at the senior level, the dominant group discriminates at the promotion stage while possibly applying meritocracy at the hiring stage. 2) Differential mingling effect: For organizational reasons, senior members tend to hang around more with senior members than with junior ones. Their homophily concerns are therefore higher for promotions than for hiring decisions.

Finally, we consider the normative implications of our analysis. To be certain, entrenchment is not always bad. First, friendship circles for example are often naturally based on homophily in tastes. Second, competition among organizations to attract talent may also imply that in situations of an exogenous supply of talent, organizations may engage in "talent-stealing" rather than promote homophily benefits. But one would expect meritocracy to be violated in many other environments. Accordingly, the paper investigates which policy interventions among mandated affirmative action, quality assessment exercises, and the overruling of majority decisions may have unintended consequences for the minority whom they are meant to benefit. For example, the external overruling of majority hiring decisions, even if justified on a stand-alone basis, may be welfare-reducing, as it leads the majority to build a larger majority cushion in order to reduce the probability that occasional interventions lead to a shift in control.

The paper then concludes by discussing alleys for future research. Omitted proofs can be found in the Appendix.

Related literature This research is related to several strands of the literature.

Discrimination theory. It shares with the literature on the economics of discrimination initiated by Becker (1957) the idea that homophily may lead organizations to disfavor minority members in their hiring decisions. Becker, though, famously emphasized that competitive market forces under some conditions make such discrimination vacuous, while we look at organizations facing imperfect market pressure. Also, Becker's analysis is static while the focus of our study is on the evolution of the organization.

In thinking about policies that protect minorities, our work is akin to the extensive literature on affirmative action (see Fryer-Loury 2005 for an overview). In Coate-Loury (1993), employers have a taste for discrimination and a principal wants to boost minority workers' incentives to invest in skills. Affirmative action gives the minority prospects and boosts minority incentives if modest, but creates a "patronizing equilibrium" and reduces incentives if extensive. In Rosen (1997)'s statistical discrimination model, a group of workers who find it hard to get a job in competition with candidates from the outgroup become less choosy; they apply for jobs for which they are less suited, and knowing this, firms rationally discriminate against group members and in favor of the outgroup.

Looking for mediocre recruits. In Carmichael (1988) and Friebel-Raith (2004), meritocracy fails because talented recruits are a nuisance for incumbent members. Carmichael argues that academic tenure makes members of a university's departments willing to hire the best possible candidates. His starting point is that an academic department coopts its new members (incumbent members of the department have better information about the potential of candidates than the administration)[2]. Tenure eliminates the incumbent members' fear of replacement by talented recruits. Similarly, Friebel and Raith consider a three-tier hierarchy (upper-level management/superior/subordinate). Because an unproductive supe-

[2]In contrast, the owner a baseball team, say, and not the players themselves, selects new recruits.

rior may in the future be replaced by a more productive subordinate, the superior deliberately hires mediocre subordinates. Hierarchical communication, which prohibits the subordinate from interacting directly with the upper-level management (and so cuts the flow of information about her talent), restores the superior's incentive to hire productive subordinates. While our model shares with these two models the view that private agendas may hinder meritocracy, the two papers involve very different modeling and study rather unrelated questions. In Mattozzi-Merlo (2015), parties pre-select candidates to build platforms and then select one of them to run in the general election. A party may pre-select a mediocre candidate over a talented one so as to preserve a level-playing field and thereby incentivize effort in the party primary.

Recruiting like-minded candidates. Our emphasis on cooptation is reminiscent of the theories of clubs (initiated by Buchanan 1965) and of local public goods (e.g. Tiebout 1956, Jehiel-Scotchmer 1997). A couple of contributions examine the dynamics of organizational membership assuming, as we do, that current members think through the impact of joiners on future recruitment decisions. They consider contexts rather different from ours, though. In particular, they stress the time variation of the size of the organization. Barberà et al (2001) look at clubs in which each member can bring on board any candidate without the assent of other members. They are interested in the forces that determine the growth or the stagnation of organizations. A member's (unilateral) decision of coopting a candidate hinges on the number of additional candidates whom the newly admitted one brings in the future; for instance, a member may not vote for his friend, because his friend may bring enemies to the group. Roberts (2015), like us, assumes majority rule, but posits that individuals care only about the (endogenous) size of the organization; there is a well-determined order of cooptation, with new members being more favorable to expansion than previous ones and therefore, if admitted, taking incumbent members into dynamics they may not wish[3]. Acemoglu et al (2012) also looks at the long-term consequences of reforms that benefit the rulers in the short run, but may imply a transfer of control in the future; for instance, a controlling elite may not want to liberalize (give political or religious rights to other citizens) by fear of a slippery slope that would later entail a loss of control.

Recruiting talent under incomplete information. Section 3.1 on homogamic evaluation capability bears resemblance with Board et al (2019). The latter paper assumes that talented people are better at identifying new talents, from which it derives rich dynamics. Section 3.1 also considers homogamic evaluation capability, but in the horizontal dimension rather than the vertical one; there may then be a separation between information and control, unlike in Board et al. Board et al also obtains virtuous and vicious spirals, but for a different reason: talented members make fewer mistakes in selecting employees, while in Section 4 of our paper, talented minority (and perhaps also talented majority) candidates turn down an organization that lacks diversity and/or talent.[4]

[3] A small literature on organizational dynamics looks at factors of hysteresis other than control over membership. In Tirole (1996) groups' reputations reflect the past behavior of their members, while members themselves have reputations based on incomplete data (that is why the individuals with whom they interact take into account the group's reputation as well). That paper shows that (uniquely determined) dynamics may converge to a high- or low- group reputation steady state, and that group reputations are fragile and hard to reconstruct once destroyed, so that a temporary shock may permanently confine a group to a low-quality trap. Sobel (2000) looks at an organization in which new recruits must "maintain the standard" of the existing population of members. He shows how, with such a rule, shocks may decrease, but not increase standards.

[4] Moldovanu and Shi (2013) model also exhibits heterogeneous evaluation capabilities. Members of a committee sequentially assessing candidates for a job and coopting using the unanimity rule each have a superior expertise in evaluating a candidate's performance along the dimension he cares most about. The focus is on the acceptance standards and the comparison between a dictator and a committee; given the focus on a single job opening, the dynamics of control are not investigated. In Egorov and Polborn (2011), similar backgrounds (homophily dimension) facilitate the estimation of others'

Trade-off between talent and like-mindedness. Cai et al (2018) analyze the dynamics of a three-member club. Like in this paper, players are characterized by a vertical and a horizontal type, and (what we label) meritocratic and entrenched equilibria may arise. Sections 2 and 3 thus generalize their analysis to an arbitrary-size organization, allowing for super-entrenchment and other types of equilibrium[5]. An interesting insight of their analysis that is not (but could be) present in our model is the possibility of "intertemporal free riding": Even in a homogenous population (which corresponds to $b = 0$ in our model), current members will not maximize social welfare; for, members in Cai et al engage in costly search for candidates. As current members are not infinitely lived and thus will not get the benefits of quality recrutement as long as the organization, they underinvest in search.[6]

Glass ceiling. In Athey et al (2000), players also have a horizontal (gender) and vertical (talent) types. Ability to fill a senior position depends on intrinsic talent and on mentoring received as a junior member. Mentoring is type-based, and so majority juniors receive more mentoring and are favored in promotions. The upper level may therefore become homogenous. The organizations however may (depending on the mentoring technology's concavity) want to bias the promotion decision in favor of minority juniors, so as to create diversity and more efficient mentoring. Control is not a focus of their paper, unlike ours.

Empirical evidence. There is growing evidence that meritocracy may not prevail even in organizations that are incentivized to behave efficiently. Zinovyeva and Bagues (2015) show that in the Spanish centralized process for promoting researchers to the ranks of full and associate professor, the promotion rate is higher when evaluated by the PhD advisor, a colleague or coauthor and that the bias dominates the informational gain (that exists with weaker connections). Bagues et al (2017) by contrast find that the presence of women on (Italian and Spanish) committees may not increase the quantity and the quality of female promotions; but male evaluators become less favorable to women if a woman joins the evaluation committee. Hoffman et al (2018) show that under discretionary hiring, the availability of test scores raises the quality of appointments (as measured by subsequent job tenure), but that the overruling of test score ranking lowers quality[7]. Rivera (2012) finds evidence of biased hiring based on shared leisure activities. Bertrand et al (2018)'s study of affirmative action on Norwegian boards (a mandated 40% female representation), together with the evidence showing that qualifications of women on boards increased rather than decreased suggests that discrimination, perhaps based on prejudice, was at stake prior to the reform[8].

2 Model

There is an infinite time horizon with periods $t \in (-\infty, +\infty)$. The organization is composed of $N = 2k$ members. At the beginning of each period, one member of the organization, drawn randomly

ability. A force pushing toward homogeneity of organizations is then the winner's curse: competition among employers makes it more likely that organizations will hire majority candidates, on whom they have superior information.

[5]Less importantly, homophily benefits are not constant in our model, while they are constant-sum (the sharing of spoils) in Cai et al.

[6]A similar effect is present in Schmeiser (2012), who analyses the dynamics of board composition and the potential benefits of outside-directors rules and nominating committee regulations. In his paper, even outside directors may not stand for shareholders' best interests, even if they can be ascertained to have no connection with insiders. The point is that, in the absence of delayed compensation, outside directors favor immediate benefits due to their limited tenure.

[7]Suggesting either homophily objectives or poor judgment.

[8]The gender gap and glass ceiling glass have a number of potential explanations, as stressed by Bertrand in her 2018 survey: difference in education (mainly in the best educational tracks), in psychological traits (higher aversion to competition/relative performance evaluation, higher risk aversion), women's demand for flexibility (particularly penalizing in professions that highly reward long hours), higher demands on time (non-market work, child penalty).

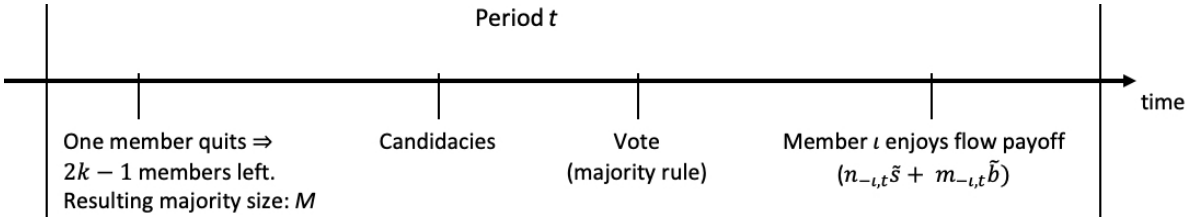

Figure 1: Timing.

from the uniform distribution, departs. We denote by δ the "life-adjusted discount factor", i.e. the pure-time discount factor times the probability of being still a member of the organization in the following period: letting $\delta_0 \in (0,1)$ denote the pure-time discount factor, then $\delta \equiv \delta_0(1-1/N)$. The departure is immediately followed by a recruitment. The intra-period timing is summarized in Figure 1.

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 $\tilde{s} > 0$ is the incremental per-period 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 per-period externality $\tilde{b} > 0$ on members of the same type, but not on members of the opposite type, and this regardless of their talent.[9,10,11]

We thus assume that each member derives utility from:

(i) colleagues' ability, i.e. the vertical attributes of members of the organization,

(ii) homophily over tastes: ceteris paribus, each member prefers colleagues who share their horizontal type.

In each period, there are two candidates for the opening, one with the same horizontal type as the organization's majority, the other with the same horizontal type as the organization's minority. Candidates apply to become members only once[12]. The candidates' types are observable prior to the vote[13]. The emergence of candidates is for the moment exogenous. The two-candidate assumption involves no loss of generality as all members of the organization always prefer the best candidate of a given horizontal type to any candidate of the same type but with lesser talent, and are indifferent if there are multiple

[9] We refer to a majority member as "he", to a minority member as "she", to a generic organization member as "they" (using the classic form of the epicene singular pronoun), and to the principal – whenever there is one – as "it".

[10] The case $\tilde{b} < 0$, corresponding to *negative homophily* – e.g. envy towards the likes, or extreme preference for diversity, etc. (see for instance Bagues and Esteve-Volart 2010) – can be easily accommodated in our model. Indeed, anticipating on the notation and our model's behavior, the set of possible flow payoffs in any period writes as $\{\tilde{s}, 0, \tilde{s}+\tilde{b}, \tilde{b}\}$. Hence, for $\tilde{b} < 0$, two cases must be distinguished:

- $\tilde{s} + \tilde{b} < 0$ (i.e. $-1 < \tilde{s}/\tilde{b} < 0$): the majority always votes for the minority candidate. The majority size converges to k, which is an absorbing state. The majority then switches and control alternates between the two groups.
- $\tilde{s} + \tilde{b} > 0$ (i.e. $\tilde{s}/\tilde{b} < -1$): the majority votes for the most talented candidate with a tie-breaking rule in favour of the minority candidate.

The simplicity of the analysis when $\tilde{b} < 0$ stems from the fact that there is then no trade-off between "quality" and "control" for the majority.

[11] Members may enjoy direct homophily benefits, associated with the desire of sharing identity (political or other) or interests (say, similar leisure activities) with fellow members. Homophily benefits may be more instrumental/ indirect. Having like-minded members on board allows one to weigh on organizational decisions and the sharing of private benefits: more committees are filled by in-group members and more suggestions favorable to the group are made. As an illustration, suppose that each member looks for a project and that search is optimally directed towards projects that favor the in-group more than the out-group (but are nonetheless rubberstamped by the out-group). Then homophily benefits are linear if projects are unrelated, and concave if there is rivalry among them (see Section 2.4 for non-linear homophily benefits).

[12] We relax this "non-storability" assumption in Section 5.1. It is made in the baseline model for the sake of exposition, as we thereby avoid the introduction of a second state variable.

[13] We relax this assumption in Section 3.1.

"best" candidates of a given horizontal type.

Let $s \equiv \tilde{s}/\left[1-\delta_0(1-2/N)\right]$ denote the expected incremental lifetime contribution of a new talented (relative to mediocre) addition to each current member of the organization[14]. We similarly denote by $b \equiv \tilde{b}/\left[1-\delta_0(1-2/N)\right]$ the expected lifetime homophily utility for an incumbent member generated by a new in-group member. So, member ι receives date-t flow payoff

$$u_{\iota,t} = n_{-\iota,t}\tilde{s} + m_{-\iota,t}\tilde{b}$$

where $n_{-\iota,t} \leq N-1$ is the number of talented colleagues and $m_{-\iota,t} \leq N-1$ is the number of in-group colleagues at date t.[15]

The decision rule is the majority rule, with each of the $2k-1$ members of the organization at the time of the vote having one vote. We denote in the following the size of (number of individuals in) the majority by $M \in \{k, k+1, ..., 2k-1\}$. We will say that the majority is tight if $M = k$.

In order to make things interesting, we assume $s > b$. Otherwise, systematically voting for the majority candidate would yield the highest possible continuation payoff for the majority, and the majority would always move toward perfect homogeneity; put differently, when $s < b$, quality considerations do not affect electoral outcomes and the majority keeps coopting majority candidates.

We let x denote the probability that the majority (or minority) candidate is more talented (i.e. has vertical type s while the other candidate has vertical type 0), and thus $(1-2x)$ is the probability that they are equally talented (either both of quality s or both of quality 0). Let α denote the probability that both are of talent s conditional on both being equally talented. Thus the probability of an in- or out-group candidate being of type s is equal to $\overline{x} \equiv x + (1-2x)\alpha$.

Our basic equilibrium concept is *perfect equilibrium in sequentially weakly undominated strategies*[16]. We rule out weakly dominated strategies so as to ignore coordination failures in which, say, a majority member votes for an unfavored candidate because other majority members also do. Concretely, each majority member votes as if he were pivotal, i.e. as if he chose the candidate.[17]

A specific subclass of equilibria restricts attention to strategies that further satisfy symmetry and Markov Perfection. Such strategies embody both symmetry (the behavior of A and B majorities are the same) and Markov Perfection (as the talents of incumbent members are no longer payoff-relevant in the sense of Maskin-Tirole 2001: a majority member's von Neumann-Morgenstern payoff function does not depend on their own talent or that of other majority or minority members). We call these *symmetric Markov Perfect Equilibria* (symmetric MPEs).

Within this latter class, we will first look for equilibria in strategies satisfying:

(i) Members of the majority (all) vote for the majority candidate if the latter is equal or superior in talent.

[14] The term $\delta_0(1-2/N)$ stems from the conditioning on both the current member and the newly recruited one still being in the organization in the next period.

[15] Alternatively we could assume that a talented member derives a "quality payoff" from her own talent, which would thus write as $\tilde{s}/(1-\delta) \neq s$. Such an assumption would leave the existence conditions unchanged, and would only marginally alter the expressions of welfare in Sections 2.2 and 6, while leaving the insights unchanged. We thus omit this possibility for notational simplicity.

[16] We refer to Acemoglu et al (2009) for a theoretical treatment of refinements in voting games. Technically, the relevant concept is their "Markov Trembling Hand Perfect Equilibrium", since the sequential elimination of weakly dominated stategies is feasible only with a finite horizon.

[17] Since we rule out coordination failures within the majority, the minority's behaviour is irrelevant (there is no absenteeism for the moment).

(ii) Members of the majority (all) vote for the majority candidate with probabilities $\{\sigma(M)\}_{M\in\{k,...,N-1\}}$ with $\sigma(M) \in \{0,1\}$, when the minority member is more talented.

We will say that the majority *switches* if it changes side. Consequently, for any symmetric MPE such that majority never switches, the majority candidate is chosen with probability 1 whenever the majority is tight (i.e. $\sigma(k) = 1$). We will say that the organization (or, equivalently, the majority) is:

- *meritocratic* if $\sigma(M) = 0$ for all M;
- *entrenched* if it favors a mediocre majority candidate over a talented minority one only when majority is tight ($M = k$), i.e. if $\sigma(k) = 1$ and $\sigma(M) = 0$ for all $M \geq k+1$;
- *entrenched at level* l if $\sigma(M) = 1$ for $M \in \{k, ..., k+l\}$, and $\sigma(M) = 0$ for $M \geq k+l+1$. Correspondingly, the organization (or the majority) is *super-entrenched* if it is entrenched at some level $l \geq 1$;
- *fully entrenched* if $\sigma(M) = 1$ for all M.

For future use, we will refer to the meritocratic and entrenched equilibria as the *canonical equilibria*.

2.1 Equilibrium characterization and existence results

Since in a Markov Perfect equilibrium, the present discounted value of benefits from other incumbent members plays no role, we do not include the legacy term in the expression of the value functions. For any group size $i \in \{1, ..., N-1\}$ just before candidacies are declared, we denote the value function of an individual in the given group by V_i: V_i is the expected discounted value of flow payoffs brought about by colleagues who will be coopted later in the period and in the future. V_i is a majority (resp. a minority) member's value function when $i \geq k$ (resp. $i < k$).[18]

We focus on the two types of equilibrium which we refer to as *"canonical equilibria"*. This specific attention will later be vindicated by Proposition 1, which establishes that *(a)* there always exists a canonical equilibrium, and that *(a)* all symmetric Markov Perfect equilibria are canonical.

Before deriving the necessary and sufficient conditions for the existence of each canonical equilibrium, we briefly investigate some properties of the value functions of majority and minority members under such strategies. Figure 2 illustrates the following lemma.

Lemma 1. ***(Properties of value functions in the meritocratic (m) and in the entrenched (e) equilibria)***

(i) (Majority value function) For $i \in \{k, ..., 2k-2\}$, V_i^e is increasing in i and has decreasing differences[19], strictly so if and only if $s > b$ and $x > 0$. Similarly, V_i^m is increasing in i and has decreasing differences, strictly so if and only if $b > 0$ and $x < 1/2$.

[18] Put differently, for any majority size $M \in \{k, ..., N-1\}$, V_M is the value function of a majority member, while V_{N-1-M} is the value function of a minority member.

[19] By "decreasing differences" (resp. "increasing differences"), we refer to the following concavity (resp. convexity) property:

$$|V_{i+1} - V_i| \leq |V_{j+1} - V_j| \qquad \Big(\text{resp. } |V_{i+1} - V_i| \geq |V_{j+1} - V_j|\Big) \qquad \text{whenever } j < i$$

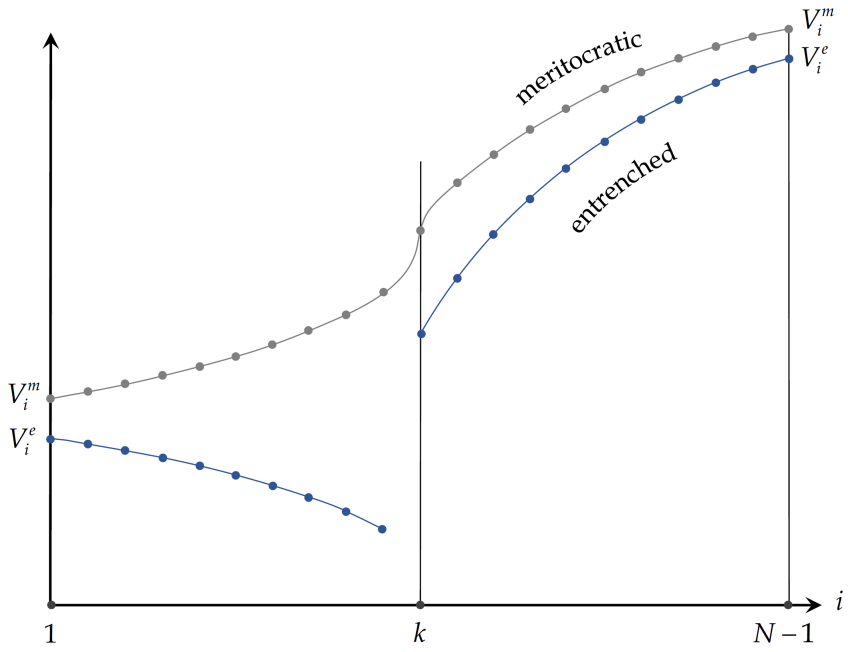

Figure 2: Properties of value functions in the meritocratic and in the entrenched equilibria.

(ii) (Minority value function) For $i \in \{1, ..., k-2\}$, V_i^e is decreasing in i and has increasing differences in i, strictly so if and only if $b > 0$, or $s > 0$ and $x > 0$. By contrast, for $i \in \{1, ..., k-1\}$, V_i^m is increasing in i and has increasing differences in i, strictly so if and only if $b > 0$ and $x < 1/2$.

(iii) (Control benefits) For $r \in \{e, m\}$ and any $i \geq k$, $V_i^r \geq V_{N-1-i}^r$, strictly so if and only if $b > 0$ and $x < 1/2$ when $r = m$, if and only if $b > 0$, or $s > 0$ and $x > 0$ when $r = e$.

Intuition. The three parts of Lemma 1 can be grasped as follows:

(i) The majority always picks its "myopically optimal" favorite candidate except in the entrenched equilibrium when $M = k$, where "myopically optimal" refers to the choice it would make in the absence of future elections or, equivalently, if future decisions did not hinge on the current one. The higher M is, the more remote the picking of a myopically suboptimal decision (entrenched equilibrium) or the loss of control (meritocratic equilibrium).

(ii) The intuition underlying the concavity/convexity of the value function for minority members is analogous to the one for majority members. The impact of moving further away from the tight-majority state fades progressively. The sign of the impact depends on the equilibrium: in the entrenched (resp. meritocratic) equilibrium, the further away from minority's size $k-1$, the smaller the additional *loss* (resp. *benefit*) of getting one step closer to $k-1$.

(iii) As discussed above, there is a benefit from control – if only because the majority members could vote like the minority members if they wanted to –. More precisely, the majority's benefit from control stems from the majority's homophily rent, i.e. the homophily payoff that accrues to majority members whenever candidates have the same talent (and which involves no loss of efficiency).

Proposition 1. ***(Canonical Equilibria)***

(i) All symmetric Markov Perfect equilibria in weakly undominated strategies are canonical.

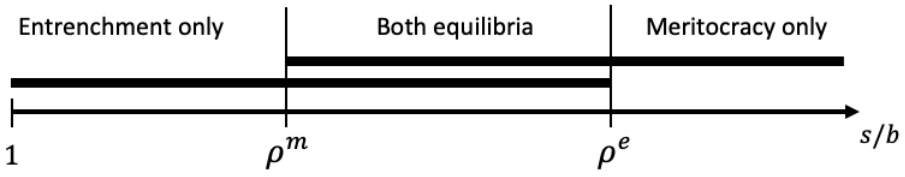

Figure 3: Existence regions for meritocratic and entrenched equilibria over the s/b line.

There exists finite thresholds ρ^e and ρ^m satisfying: $1 \leq \rho^m < \rho^e < +\infty$, such that

(ii) The entrenched equilibrium exists if and only if $s/b \leq \rho^e$,

(iii) The meritocratic equilibrium exists if and only if $s/b \geq \rho^m$.

(iv) Patience fosters entrenchment: for any δ, $\partial\rho^m/\partial\delta \geq 0$, and $\partial\rho^e/\partial\delta \geq 0$.

Figure 3 describes the existence regions over the line s/b for given x, δ. For s/b close to 1, there is little cost for the majority to select an untalented peer over a better qualified minority candidate. But there is a benefit from keeping control: whenever both candidates have the same ability, majority and minority voters differ on the candidate they would like to recruit, regardless of any control benefit. Hence, being in the majority allows to pick the majority candidate, yielding a homophily benefit and no cost in terms of quality. And so the majority is entrenched.

As the ratio quality/homophily payoffs s/b increases, the cooptation game moves from a (bounded) region where only the entrenched equilibrium exists, to an intermediate (bounded) interval where both equilibria coexist. As s/b continues to increase, it reaches the (half-line) region where only the meritocratic equilibrium exists.

As the discount factor increases, the existence region of the meritocratic equilibrium shrinks while that of the entrenched equilibrium widens. These comparative statics are intuitive as when members become more patient, the cost of losing the majority to the outgroup increases.

Remark. If $x = 1/2$, i.e. the probability that both candidates have the same vertical type is nil, then $\rho^{\mathrm{m}} = 1$: for any $s \geq b$, there exists a meritocratic equilibrium. The result is intuitive, as there is no pure benefit from control.

2.2 Welfare

2.2.1 Non-ergodic welfare.

We first consider current members' welfare, defined as their total surplus (generated by both quality and homophily), at any given legacy and period. We refer to this welfare notion as "non-ergodic welfare". As it turns out, the meritocratic equilibrium is preferred to the entrenched equilibrium by all members of the organization. At any given majority size, minority members obviously prefer the meritocratic equilibrium, while majority members, who can always select to be entrenched, weakly prefer the meritocratic equilibrium which delivers a higher payoff when surrendering control.

Proposition 2. ***(Non-ergodic welfare)*** *Whenever the meritocratic and the entrenched MPE coexist, i.e. for $s/b \in (\rho^m, \rho^e)$, at any majority size the meritocratic equilibrium is preferred by all members of the organization to the entrenched equilibrium.*

2.2.2 Ergodic aggregate welfare.

We now draw an aggregate-welfare comparison between entrenchment and meritocracy in their respective ergodic distribution from the ("objective") perspective of a principal or third-party putting at least as much weight on quality as on homophily benefits. We first describe the ergodic distributions of majority sizes. Since payoffs in a given period accrue after the current-period vote and before the next-period departure, we are interested in the *end-of-period* distribution of majority sizes. Index the end-of-period majority size by $i \in \{k, ..., N\}$. Let ν_i^{r} denote the ergodic probability of state i at the end of a period in regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$ (see Appendix E for their expressions).

Lemma 2. ***(End-of-period ergodic distributions)*** *The probability distribution $\{\nu_i^e\}$ strictly first-order stochastically dominates $\{\nu_i^m\}$.*

Ergodic quality. By taking the fixed point of the dynamic equation for (expected) aggregate quality in the ergodic state[20], one has

$$\begin{cases} S^{\mathrm{m}} & \equiv N(N-1)\big(\overline{x}+x\big)\tilde{s} \\ S^{\mathrm{e}} & \equiv N(N-1)\left[\nu_{k+1}^{\mathrm{e}}\dfrac{k+1}{N}\overline{x}+\left(1-\nu_{k+1}^{\mathrm{e}}\dfrac{k+1}{N}\right)\big(\overline{x}+x\big)\right]\tilde{s} \end{cases}$$

Unsurprisingly, the ergodic efficiency of a meritocratic organization always exceeds that of an entrenched one:

$$S^{\mathrm{m}} - S^{\mathrm{e}} = N(N-1)\nu_{k+1}^{\mathrm{e}}\frac{k+1}{N}x\tilde{s} > 0.$$

Ergodic homophily benefit. For regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$, the aggregate per-period homophily benefit writes

$$B^{\mathrm{r}} \equiv \sum_{i=k}^{N} \nu_i^{\mathrm{r}}\Big[i(i-1)+(N-i)(N-i-1)\Big]\tilde{b}$$

An entrenched organization always dominates a meritocratic one in terms of ergodic aggregate homophily benefit ($B^{\mathrm{m}} < B^{\mathrm{e}}$): *(a)* the function $\big(i \mapsto i(i-1)+(N-i)(N-i-1)\big)$ is strictly increasing for $i \in \{k, ..., 2k\}$, and *(b)* the probability distribution $\{\nu_i^{\mathrm{e}}\}$ strictly first-order stochastically dominates $\{\nu_i^{\mathrm{m}}\}$ from Lemma 2.

Ergodic aggregate welfare. Define the ergodic per-period aggregate welfare in regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$ as $W^{\mathrm{r}} \equiv qS^{\mathrm{r}} + B^{\mathrm{r}}$, with $q \geq 1$, allowing for external spillovers of the organization's quality.

Proposition 3. ***(Ergodic per-period aggregate welfare)*** *For any $s > b$, $W^m > W^e$, i.e. the meritocratic equilibrium dominates the entrenchment equilibrium in terms of ergodic per-period aggregate*

[20] The aggregate quality at the end of period $t+1$ is the aggregate quality at the end of period t minus the (expected) loss due to a member's departure, plus the (expected) contribution of the recruited candidate. For the meritocratic equilibrium,

$$S_{t+1}^{\mathrm{m}} = \frac{N-1}{N}S_t^{\mathrm{m}} + (N-1)[\overline{x}+x]\tilde{s}$$

Similarly for the entrenched equilibrium,

$$S_{t+1}^{\mathrm{e}} = \frac{N-1}{N}S_t^{\mathrm{e}} + (N-1)\left[\nu_{k+1}^{\mathrm{e}}\frac{k+1}{N}\overline{x}+\left(1-\nu_{k+1}^{\mathrm{e}}\frac{k+1}{N}\right)\big[\overline{x}+x\big]\right]\tilde{s}$$

welfare, when the latter puts at least as much weight on quality as on homogamy.

2.3 A continuum of vertical types

We have assumed so far that talent can take only two values. We show in this section that our previous insights still hold when talent is continuously distributed within a class of equilibria: (i) a stronger majority engages in more meritocratic recruitments, and (ii) whenever several equilibria coexist, they can be ranked from more to less meritocratic, and Pareto-compared.

We look for equilibria that can be described as a sequence of cut-offs $(\Delta_M)_{M\in\{k,...,N-1\}}$ such that whenever a majority of size M recruits the out-group candidate with (discounted) talent $\hat{s}$ against the in-group candidate with (discounted) talent s if and only if $\hat{s}-s>\Delta_M$ where $\Delta_M>b$ (so the question now is not whether there is discrimination, but how much discrimination there is). Lastly, we denote by $\prec$ the order relation defined over the set of decision rules such that $\Delta\prec\Delta'$ if and only if $\Delta_M<\Delta'_M$ for all $M\in\{k,...,N-1\}$. We will then say that the former decision rule is more meritocratic.

Let $\mathcal{G}$ be the set of continuous joint distributions of $(s,\hat{s})$, i.e. resp. the quality of the majority and the minority candidate, with support[21] $[0,+\infty)^2$ such that $\mathbb{E}[\max(\hat{s},s+b)]<\infty$, and $(\hat{s}-s)$ is symmetrically distributed around 0 with $\mathbb{P}(\hat{s}-s>b)>0$ and such that, letting the function h be defined by

$$h(\Delta)\equiv\mathbb{E}\big[(s+\Delta)\mathbf{1}\{\hat{s}-s\leq\Delta\}\big]+\mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s>\Delta\}\big],$$

the functions $[h(\Delta)-\Delta/2]$ and $[\Delta-h(\Delta)]$ are strictly increasing with $\Delta\in(b,\overline{s})$ where $\overline{s}=\sup(\hat{s}-s)$. This set includes the set of (full support) continuous joint symmetric distributions. It also includes the case where the majority candidate has a fixed type $s\geq 0$ and the minority candidate a type $s+D$ where D is a (full support) random variable with a continuously differentiable distribution over $(-s,s)$ symmetric around 0.

Proposition 4. ***(A continuum of vertical types)*** *Assume talent is distributed according to a joint distribution $G\in\mathcal{G}$. Then there exists a non-empty class of equilibria such that the sequence $(\Delta_M)_M$ is strictly decreasing: a stronger majority discriminates less than a weaker majority. Moreover, any two equilibria within this class, with distinct decision rules Δ and Δ', can be ranked by the order relation $\prec$. If $\Delta\prec\Delta'$, then the equilibrium characterized by the decision rule Δ (which is more meritocratic than the one described by Δ') is preferred, for δ small, at any majority size by all members of the organization to the latter.*[22,23]

2.4 Non-linear homophily benefit

A non-linear homophily benefit does not require enlarging the state space, as the size of the majority is still a sufficient statistics looking forward. While the homophily benefit of an extra in-group member

[21] We show the result more generally (i.e. also for distributions with finite support) in Appendix G.

[22] We further show in the Appendix that for any δ, any "meritocratic" equilibrium (i.e. with $\Delta_k<\overline{s}$) is preferred at any majority size by all majority members to the entrenched equilibrium ($\Delta_k=\overline{s}$), if these coexist.

[23] As a consequence, given a joint distribution $G\in\mathcal{G}$, in any equilibrium within this class, whenever the majority is not tight, it recruits a minority candidate with a strictly positive probability. In addition, we show that for distributions such that $\mathbb{P}(\hat{s}-s>\Delta)>0$ for any $\Delta<\infty$, in any equilibrium within this class, the majority recruits a minority candidate with a strictly positive probability at any majority size. Hence control switches happen with strictly positive probability.

depends on future hirings under a non-linear homophily benefit, the key trade-offs (driven by meritocracy vs. control) are not affected.

Let $\tilde{\mathcal{B}}(i)$ denote the per-period homophily benefit enjoyed by a member whose in-group has size i (thus, in the linear case, $\tilde{\mathcal{B}}(i) \equiv (i-1)\tilde{b}$).

(a) Concave homophily benefit. Suppose, first, that $\tilde{\mathcal{B}}(i)$ is concave in the number of in-group members i, and that $\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k) < \tilde{s}$ (otherwise super-entrenchment obtains[24]). The same analysis as in Section 2.1 shows that the equilibrium is either meritocratic or entrenched.

The concavity of $\tilde{\mathcal{B}}(i)$ may favor entrenchment or meritocracy. If $\tilde{\mathcal{B}}(i) \equiv (i-1)\tilde{b}$ for $i \geq k$, then concavity favors entrenchment, as the payoff under entrenchment is the same as in the fully-linear-homophily-benefit case, while the payoff under meritocracy is smaller. Symmetrically, if $\tilde{\mathcal{B}}(i) \equiv (i-1)\tilde{b}$ for $i \leq k$, the benefits of entrenchment are smaller under concavity while losing control has identical costs. Overall, concavity has ambiguous effects on the prevalent equilibrium.

(b) Convex homophily benefit. The analysis requires some adaptation in the case of convex homophily benefits[25]. We do not offer a full analysis, and content ourselves with the following observation. Suppose that the homophily benefit is linear up to $i = N-1$, but a large payoff accrues from full homogeneity (so that $\tilde{\mathcal{B}}(N) - \tilde{\mathcal{B}}(N-1)$ is larger than $\tilde{s}$). Then the organization may be meritocratic for small majorities and no longer so for large ones: the majority's expected cost of building a full majority (and maintaining it thereafter) becomes smaller as the majority size increases. While the ergodic state exhibits full entrenchment, the dynamics differ from the other instances of full entrenchment exhibited in the paper and may be meritocratic for a while.

Observation 1. The analysis carries over to concave homophily benefits. By contrast, convex homophily benefits may give rise to new organizational dynamics.

3 Super-entrenchment

The most obvious case for super-entrenchment is $s \leq b$, which indeed leads to full entrenchment. We just noted that concave homophily benefits may lead to super-entrenchment. Section 6 will show that some well-meaning interventions may have the unintended consequence of incentivizing the majority to be super-entrenched. Besides these three reasons, four other drivers of super- and full-entrenchment are studied in this section.

3.1 Homogamic evaluation capability

We have assumed so far that all members are equally proficient at evaluating the talents of in- and out-group candidates. However some environments exhibit an asymmetry in this ability. For example, econometricians are better placed than development economists to evaluate an econometrician, and con-

[24] Namely, if there exists $l \in \{1, ..., k-1\}$ such that

$$\tilde{\mathcal{B}}(k+l) - \tilde{\mathcal{B}}(k+l-1) > \tilde{s} > \tilde{\mathcal{B}}(k+l+1) - \tilde{\mathcal{B}}(k+l),$$

then the only equilibrium is super-entrenchment at level l. Indeed, the myopically optimal choice allows the majority to keep control. Hence it can guarantee itself the upper bound on its payoff.

[25] Convexity may arise for instance when facilities or regulations must be added to accommodate the existence of a minority, or if a group's reaching a critical size delivers additional opportunities to its members.

versely. This section investigates how the analysis is affected when only in-group evaluation is feasible.[26]

The majority still selects the majority candidate if the latter has quality s. So we can focus on the situation in which the majority candidate has quality 0. The conditional quality of the minority candidate is then

$$s^\dagger \equiv \frac{x}{x+(1-2x)(1-\alpha)}s = \frac{x}{1-\overline{x}}s$$

Two mutually exclusive cases must be distinguished.

Case 1: $b \geq s^\dagger$. This case arises when correlation is high (x low) and average quality low ($\overline{x}$ low), so the majority is pessimistic about the minority candidate's talent when its own candidate lacks talent. [Departing from the Bayesian framework, case 1 would also be more likely if the majority members had a rather negative stereotype about minority members' talent.]

When $b \geq s^\dagger$, the majority is fully entrenched: it keeps admitting solely majority candidates and ends up being homogeneous. This implies that imperfect information (in the form of homogamic evaluation capability) may transform an entrenched or meritocratic organization into a fully entrenched one.

Case 2: $b < s^\dagger$. For case 2 to arise, majority members need to be sufficiently optimistic about the average quality of minority candidates. That is, the draws in talent must be sufficiently uncorrelated (i.e. x large) and the average ability of a candidate high enough (i.e. $\overline{x}$ large). [Had we assumed non-Bayesian beliefs, a further condition would have been the absence of prejudice about the minority.]

We provide intuition for the results before starting the analysis. When $b < s^\dagger$, the model becomes similar to our baseline setup, yet with two crucial changes:

(i) The probability that the minority candidate is assessed by majority members as strictly more talented (in expectation) than the majority one increases from x to $x^\dagger \equiv x + (1-2x)(1-\alpha) > x$. In other words, minority candidates may get the benefit of the doubt.

(ii) The stand-alone cost of an entrenched vote is smaller as $s^\dagger - b < s - b$.

We show that, except perhaps when the majority is tight ($M = k$), whenever the majority candidate lacks talent, the majority gives the benefit of the doubt to, and picks the minority candidate. This means that the minority candidate may be selected even though the two candidates are equally talented. The majority candidate is selected with probability $1 - x^\dagger$ and the minority candidate is selected with probability $x^\dagger$. The homogamic choice has probability below 1/2 if and only if $\alpha < 1/2$. Even more strikingly, for $\alpha < 1/2$, the majority's choice then makes the minority happier about the choice than the majority itself: the expected quality benefit from the appointment is the same for both groups, while the homophily benefit is $(1-x^\dagger)b$ for a majority member and $x^\dagger b > (1-x^\dagger)b$ for a minority member. *Hence, there is a curse of control*[27]. We show that for $\alpha < 1/2$ the two canonical equilibria still obtain: as is intuitive, the meritocratic one then exists for any $s^\dagger > b$, while the entrenchment one exists in a bounded region. Indeed, while the existence of the latter might seem surprising, it results from the following

[26] Our analysis in this section is related to literature on asymmetric evaluation capability – see for instance Moldovanu-Shi (2013) –, although to our knowledge, our cooptation-oriented approach is new.

[27] This effect may depend on our modelling assumptions. When the majority candidate lacks talent, majority members would presumably prefer to postpone the recruitment, de facto deneging the benefit of the doubt to the minority candidate. Our model rules out this option by assuming a recruitement must be made in each period – e.g. because it is too costly for the organization to go under-staffed for one period.

trade-off: although the minority benefits more from the new recruit than the majority whenever the majority size is not tight, the opposite holds when $M = k$, which happens frequently since $x^\dagger$ is high.

The same arguments as with perfect information apply, with the appropriate changes in payoffs and with $x^\dagger$ replacing x in the transition probabilities. We focus on the following two equilibria which are the analogs of the perfect-information canonical equilibria.[28]

Proposition 5. ***(Canonical equilibria with homogamic evaluation capability)***

(i) If $b \geq s^\dagger$, the majority coopts only candidates of the in-group and therefore becomes homogeneous.

(ii) If $b < s^\dagger$ (i.e. $s/b > x^\dagger/x$), there exists an equilibrium in which for all $M \geq k+1$, the majority votes for the majority's candidate if talented, and for the minority's candidate (of unknown talent) otherwise. There exist finite thresholds $\rho^{e\dagger}$ and $\rho^{m\dagger}$ satisfying[29]

- *The entrenched equilibrium (in which the majority always chooses the majority candidate for $M = k$) exists if and only if $s/b \leq \rho^{e\dagger}$.*
- *The meritocratic equilibrium (in which the minority candidate is elected against an untalented majority candidate even for $M = k$) exists if and only if $s/b \geq \rho^{m\dagger}$.*

Remark: Cheap talk. One may wonder whether communication could help the majority select a candidate. The answer is that, for $x^\dagger \leq 1/2$, cheap talk cannot operate in this environment due to a form of winner's curse. Because the majority picks its candidate whenever talented, the minority infers that whatever message it sends can only have an impact when the majority candidate is untalented. Conditional on a low-quality majority candidate, the minority always prefers its own candidate, and so any message sent to the majority is necessarily uninformative.

Remark: Intermediate assessment abilities. We have so far assumed that a group is able to access the quality of outgroup members either perfectly or not at all. Intermediate assessment abilities would give rise to additional and interesting insights. One might imagine in particular that having more minority members in the organization brings more familiarity with their characteristics and therefore an enhanced ability to assess outgroup candidates' ability. We conjecture that dynamics similar to those of Section 4 for endogenous candidacies would then arise: the majority may then want to voluntarily engage in (limited) affirmative action for "talent intelligence" purposes; virtuous and vicious circles would similarly emerge.

3.2 Uncertain voting participation and absenteeism

We have assumed so far that all members of the organization vote. Absenteeism, whether due to illness or alternative obligations, may incentivize the majority to secure majorities of more than one vote so as to minimize the probability of a majority switch. Even large polities may find it optimal to stand in the way of talented minority candidates.

Returning to symmetric evaluation capability, we first model absenteeism in a general fashion before providing an explicit illustration. Namely, for any majority size $M \in \{k, ..., N-1\}$, let $\Lambda(M)$ be the probability that, because of absenteeism, a majority of size M loses the vote, i.e. that the minority's

[28] As with perfect information, our equilibrium concept rules out coordination failures within the majority, and thus the minority's behaviour becomes irrelevant.

[29] If $b < s^\dagger$ and $x^\dagger \geq 1/2$, then $\rho^{m\dagger} \leq x^\dagger/x$, and thus the meritocratic equilibrium exists for all $s/b \geq x^\dagger/x$.

opinion prevails[30]. We assume that the majority is strictly more likely than the minority to win the vote, and the more so, the greater the majority size, and is certain to win for sufficiently large majority sizes (perhaps $N-1$)[31]:

$$\begin{cases} \Lambda \text{ decreases with respect to majority size } M \\ \Lambda(M) \in (0,1/2) \quad \text{for any } M \in \{k, ..., k+l-1\}, \quad \text{and} \quad \Lambda(M) = 0 \quad \text{for any } M \geq k+l \end{cases} \tag{1}$$

While the Λ function can capture correlation in absenteeism, either within groups or across the entire population of members, an interesting case occurs when absences are i.i.d. (the Bernoulli case). That case satisfies (1) with $\Lambda(M) > 0$ for all $M < N-1$. While we allow for a wide range of absenteeism functions (in particular as we allow for correlation in voting turnout), condition (1) may not always be warranted if voting participation is strategic rather than caused by exogenous events.

We look for monotonic[32] symmetric MPEs in weakly undominated strategies, which indeed exist.

Remark. In contrast to the baseline model, the minority's strategy now matters at any majority size. Because the minority's probability of being pivotal is positive for $M \leq k+l-1$, it is in fact an equilibrium requirement for minority members to behave as if they picked the outcome.

Proposition 6. ***(Absenteeism and super-entrenchment)*** *Let Λ satisfy (1) and $x < 1/2$. For s/b sufficiently close to 1, super-entrenchment at level l is the unique symmetric MPE in weakly undominated strategies such that a stronger majority makes (weakly) more meritocratic recruitments. The minority's equilibrium strategy consists in always voting for its own candidate at any majority size $M \leq k+l-1$. Furthermore, for s/b sufficiently close to 1, in any symmetric MPE in weakly undominated strategies, the majority is entrenched when it has size $k+l$.*

In particular, if $l = k-1$ as in the Bernoulli case, the possibility of absenteeism may trigger full-entrenchment for any s/b sufficiently close to 1.

When Λ satisfies (1) with $l < k-1$, the majority is "safe" at any majority size $M \geq k+l$ as it still controls the outcome with probability 1. Therefore, meritocracy, i.e. picking the minority candidate whenever she is strictly more talented, is optimal at these majority sizes.

3.3 Other drivers of the size of entrenchment

The model captures the private and social costs and benefits of entrenchment. The strong ability of the majority to control majority switches underestimates, for at least two more reasons, the extent to which entrenched organizations actually keep talented minority candidates at bay.[33]

(a) *Imperfect identification of group allegiance.* We introduce the possibility that a candidate be able

[30] We assume that absenteeism in a given period is independent of the candidates' qualities in that given period: in particular, absenteeism does not result from members' strategic decisions given candidates' types.

[31] Absenteeism raises the question of what happens when the numbers of majority and minority members who show up are equal (or if no-one shows up). The key assumption behind the statement of the Λ function is that a process is in place, which will guarantee a decision in case of such draws. One can envision a variety of such processes. For example, the majority leader might take the decision. Or the assembly of members might reconvene as many times as is needed to break the tie (technically, an infinite number of times if one wants to reach a decision with probability 1. Otherwise the results are just limit results). Similarly, one could add a quorum rule given such reconvening; this quorum, for a given absenteeism process, would generate a different Λ function, but still one satisfying our assumptions. The Λ function captures all kinds of processes and all forms of correlation among members' absences, as long as the process delivers an outcome.

[32] In the sense that a stronger majority makes more meritocratic recruitments.

[33] Moreover, if vertical types are continuously distributed (see Section 2.3), then in any of the equilibria of Proposition 4, the organization is never fully meritocratic.

to mascarade as belonging to the other group and thereby be elected. Namely, we assume there is a probability $\vartheta \in (0, 1/2)$ that the best candidate of the majority group[34] is incorrectly identified (tagged as belonging to majority group, when actually belonging to the minority group). To avoid having to consider complicated disclosure strategies of misidentified members, we further assume that the real identity of the newly elected member is revealed after the vote and before curent-period payoffs accrue.

The probability of a fully-entrenched majority with size $M = N - 1$ losing control, is strictly positive and proportional to ϑ^k, as it takes k consecutive occurrences of "bad luck" to topple its grip on the organization. By the usual argument[35], there exists a non-empty neighbourhood of 1 such that for s/b in this neighbourhood, the (only monotone) equilibrium is the fully-entrenched equilibrium.

This analysis of turncoats presumes that candidates identified as sympathetic to the majority may actually favor the minority. A milder version of the same idea is that candidates identified as pro-majority may actually prefer a majority candidate, but with an intensity that is not observable at the moment of their election. So a majority recruit may put more weight on talent relative to homophily than the average majority member[36] and therefore resist the entrenched strategy. Anticipating this possibility, the majority might again want to be super-entrenched, so as to minimize the probability of a switch in control.

(b) *Supermajority clause for some decisions.* The case in which each period, a non-hiring decision is subject to a supermajority is similar to (locally) convex homophily benefits. We illustrate this by considering unanimity. Assume the decision yields $\tilde{b}^+$ for majority members where $\tilde{b}^+ + \tilde{b}$ is significantly larger than $\tilde{s}$ (and maybe yields something very negative for minority members to justify the rule). Then the setting is similar to the example we gave for convex homophily benefits.

Observation 2. The imperfect identification of group allegiance, and the existence of supermajority requirements for some non-hiring decisions may both give rise to super- or full-entrenchment.

4 Endogenous candidacies: Voluntary affirmative action and virtuous/vicious spirals

Organizations may not be able to hire talents who have attractive outside options. This section first fully characterizes organizational dynamics when outside options are exogenous, and then obtains partial results when organizations compete. The key insights are (a) the emergence of vicious spirals in which talented minority members, and then possibly talented majority members turn down offers, and (b) the possibility that the organization *voluntarily* adopts affirmative action so as to later make itself more attractive to talented minority candidates.

[34] We implicitly assume that all candidates of the majority group are equally "unreliable" (incorrectly identified with the same probability). Alternatively, a richer modelling would allow for heterogeneity within a group: an untalented yet fully "reliable" candidate (i.e. identified as perfectly belonging to the majority) may then be preferred to talented yet "unreliable" candidates.

[35] The same argument as in Section 3.2 applies, with the probability of the majority losing the vote becoming the probability of recruiting a minority candidate incorrectly identified.

[36] For example, a small fraction of majority candidates might have homophily benefit zb, where $z < 1$, and a preference for the meritocratic strategy over the entrenched one favored by their colleagues in the majority.

4.1 Exogenous outside options

We endogenize candidacies. In order to keep the model tractable, we focus on large organizations and furthermore study the continuous-time limit of the discrete model[37]. When candidates have outside options, their decision to apply to, or to accept joining the organization is both forward- and backward-looking. It depends on the identity of the majority group, the size of its majority and the average quality of incumbent members. The complexity added by the legacy quality's relevance leads us to focus on a large organization, for which the equilibrium can be described through a phase diagram.

The organization has a unit mass of members. Time is continuous. 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 of untalented candidates of each group, as well as $x\chi dt$ talented candidates from each group, where $x < 1/2$. Candidates 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 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.

Candidates' participation decisions are intertemporal strategic complements. 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.

Letting $M \in [1/2, 1]$ denote the majority's size and S the fraction of talented members (so that the current quality of the organization is equal to $S\tilde{s}$), the flow payoff of a minority member is

$$S\tilde{s} + (1-M)\tilde{b}$$

Let σ_1 (resp. σ_2) denote the 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 in-group candidates over equally untalented out-group ones).

First, note that in large organizations the majority is freed from the vagaries of a random pool of candidates, and, due to symmetry, never faces a tradeoff between sacrificing quality and losing control[38].

[37] For the sake of consistency, we investigated the discrete model for an arbitrary size, under a minimal information assumption; the material is available upon request from the authors.

[38] The main drawback of this deterministic model of large organizations is that control is no longer an issue (indeed, meritocracy allows the majority to keep control in the bare-bones version of the model). This feature is inconsequential for the investigations of various dynamics under entrenchment, as is the case here. Furthermore, one can reintroduce control concerns in the large-organization model by adding persistent shocks. For example, the relative absenteeism of the majority vs. minority members might follow a Brownian motion.

Another (but asymmetric) case would be one in which there are more talented B-candidates than talented A-candidates: $x_A < x_B$. If $x_B > 1/2$ and if there is still 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

We look for an equilibrium such that the majority solves an optimal control problem, without having to worry about the possibility of losing control[39]. Given that control is no longer a consideration in formulating strategies, both the majority and the minority prefer to take talented candidates as long as the extra flow payoff from a talented candidate exceeds the homophily benefit, which we keep assuming, and so $\sigma_1 = \sigma_2 = 1$ whenever the participation constraint of each type is met.

Lastly, we define the expected intertemporal utilities: $s \equiv \tilde{s}/(r+2\chi)$, $b \equiv \tilde{b}/(r+2\chi)$ and $u \equiv \tilde{u}/(r+2\chi)$.

We make the following assumptions:

(The organization may attract minority talents) Under parity and meritocracy, talented members receive a positive net surplus:

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

(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:

$$2xs + xb < u \tag{3}$$

These two assumptions together will later guarantee the existence of an interior steady state with majority size:

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

In the region of the parameter space in which talented minority candidates accept to become members (regions 1 and 2 below), the flow-quality dynamics are given by

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

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:

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

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 $s/b \geq (1/2 - x_A)\big/(x_B - 1/2)$.

[39]Namely, 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.

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 surplus in the organization), the majority favors its own candidates in the untalented group ($\sigma_0 = 1$):

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

- 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:

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

  The assumption on the viability of a meritocratic organization implies that $\sigma_0^* > 0$, while the assumption that talented minority candidates' outside option constrains the majority implies that $\sigma_0^* \leq 1$. Whenever the organization reaches region 2, it monotonically converges to the steady state (S^*, M^*), which lies in region 2.[40]

- 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

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

- 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

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

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

Figure 4 depicts the phase diagram of the organization's current quality S and majority size M when $u \in [2xs + xb, 2xs + b/2]$ and $u < \max(xs + (1-x)b, b/2, 3xs)$.[41]

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:
(Affirmative action may be attractive) The majority's flow payoff from newcomers is higher in the high-

[40] 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.

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

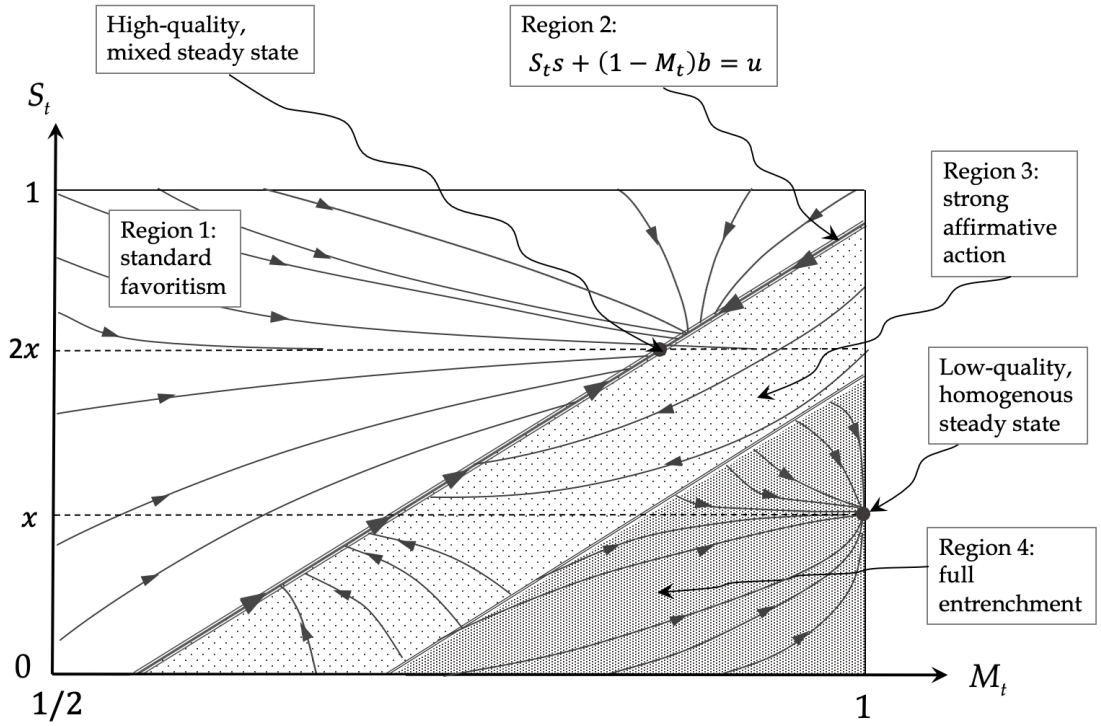

Figure 4: Phase diagram of the organization's current quality S and majority size M when $u \in [2xs + xb, 2xs + b/2]$ and $u < \max(xs + (1-x)b, b/2, 3xs)$.

quality steady state than in the low-quality one:

$$2xs + xb + (1 - 2x)\sigma_0^* b > xs + b, \qquad \text{i.e.} \qquad 3xs > u, \tag{4}$$

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). 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 region 2, which gives them a zero net surplus, their net utility of joining overall is strictly negative.

We show in Appendix J.1 that region 3 is non-empty if and only if the following condition holds: *(Affirmative action can lure back talented minority candidates)*[42,43]

$$xs + (1 - x)b - u > 0 \tag{5}$$

If (5) 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*

[42]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. $xs + (1-x)b - u > 0$.

[43]The assumption that $u \in [2xs + xb, 2xs + b/2]$ combined with (5) implies in particular that $x < 1/3$.

sacrifice is worth the trouble starting at (M_0, S_0): for $S_0 s - M_0 b < u - b$,[44]

$$\int_0^T e^{-(r+\chi)t}(1-x)b\big[1-e^{-\chi t}\big]dt + \int_T^{\infty} e^{-(r+\chi)t}(1-x)b\big[1-e^{-\chi T}\big]e^{-\chi(t-T)}dt$$
$$\leq \int_T^{+\infty} e^{-(r+\chi)t}\big(3xs-u\big)\big[1-e^{-\chi(t-T)}\big]dt \tag{6}$$

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 (6) may thus be rewritten as: $\underline{u} < S_0 s + (1-M_0)b < u$ for some $\underline{u} > 0$.

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. 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 Appendix, and summarize the key messages in Proposition 7.

Proposition 7. ***(Endogenous candidacies, voluntary affirmative action and virtuous/vicious spirals in large organizations)*** *Assume (2)-(3). There exists an MPE satisfying:*

(i) Equilibrium uniqueness and steady states: There exist at least two steady states if talented candidates have an outside option: $(M^*, S^*) \in \{(x+(1-2x)\sigma_0^*, 2x), (1,x)\}$. *A third steady state,* $(1,0)$*, may exist. Both majority and minority members rank the steady state* $(x+(1-2x)\sigma_0^*, 2x)$ *first,* $(1,x)$ *second and* $(1,0)$ *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, 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)*[45]*. 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*

[44] As the LHS in (6) is strictly positive for $T = 0$ (from assumption (4)), condition (6) 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 (6) strictly increases with T while the RHS strictly decreases with T, if condition (6) 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 (6) holds with equality. Hence the boundary between Regions 3 and 4 is an (increasing) line in the plane (M, S).

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

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).*[46]

(iii) Voluntary affirmative action: There exists a range of initial states (region 3)[47] *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 $(M^*, S^*) = (1, x)$ *or* $(1, 0)$ *(the latter if and only if the initial state lies in region 5 and* $u > b$*).*

(v) Who drops out first when quality decays: In the long run, as the initial quality S_0 *decreases, talented minority candidates reject the organization's offers before talented majority candidates do.*

Remark: Quits. If the organization's talented current members also have access to the same outside option, then the organization may lose all its talented minority members at once. Intuitively, this would put an additional constraint on the profitability of engaging in affirmative action, thus reducing 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 S_{t^+} depend on the initial distribution of quality *within each group.*

4.2 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 (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[48]. 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

[46] 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.

[47] Region 3 is non-empty if and only if (4)-(5) hold.

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

group-coalition proof[49].

Assume organization 1 starts with a higher quality ($S_2(0) < S_1(0) \leq 2x$)[50] and, say, majoritarian group A, while organization 2 starts with a B majority[51]. 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:

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

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):[52]

$$\big[S_1(0) - S_2(0)\big]s + \big[\big(1 - M_1(0)\big) - M_2(0)\big]b \geq \frac{\chi}{r+\chi}\big[(1-x)b - xs\big] \tag{7}$$

Pursuing the analysis along these lines, we can show:

Proposition 8. *(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,*
- *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 Anterooms for appointments

We have so far viewed the appointment process as an organizational choice between coopting candidates and letting them go away for good. While a first step, this assumption ignores the possibility that appointments may result from a dynamic process operating outside or inside the organization. First, turned-away candidates may be persistent and later reapply. Second, the organization may groom ju-

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

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

[51] Alternatively, insights are unaltered if it starts with an A majority (see Appendix K).

[52] The group-coalition proofness condition for talented A-group candidates is a fortiori satisfied when (7) is:

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

nior members for possible promotion to senior positions. This section analyzes in sequence these two possibilities, which display several similarities.

5.1 Candidates can re-apply

We investigate the consequences of unselected candidates being able to re-apply. "Stored" candidates keep re-applying until they are recruited[53]. For the sake of exposition, we make a further simplyfing assumption: $\alpha = 0$, so that in any period, the new majority and the minority candidates are equally talented if and only if they both are untalented (which happens with probability $(1 - 2x)$), and the unconditional probability that a new candidate is talented is given by $\overline{x} = x$. This assumption implies that under meritocratic hiring, talented candidates are always hired and so the ability to re-apply is irrelevant on an equilibrium path.

Proposition 9. ***(Reapplying for membership)*** *Assume $\alpha = 0$. Entrenchment yields the majority a higher value when candidates reapply than when they cannot: being able to "keep in store" a talented minority candidate when the majority is tight reduces the cost for the majority of turning down her application. Moreover, the existence region for the meritocratic equilibrium shrinks when the organization can store applications.*

5.2 Hierarchies and the glass ceiling

The expression "glass ceiling" refers to the difficulty for women (or minorities) to rise beyond a certain level in a hierarchy. While there are various hypotheses for its existence, whose relevance is reviewed in Bertrand (2018), we here investigate whether non-meritocratic cooptation might be a factor. Even if male dominance and favoritism contribute to discrimination against women, it is not a priori obvious that they imply a lower rate of promotion for women and therefore a glass ceiling.

Consider a large two-tier organization with a mass 1 of senior positions and a mass $J > 1$ of junior positions. A higher J corresponds to a "more pyramidal" organization. Between times t and $t + dt$, a fraction $\chi^S dt$ of seniors departs and is replaced by juniors promoted to seniority; a fraction $\chi^J J dt$ of juniors departs as well. To offset these two flows out of the junior pool, a fraction $\hat{\chi} J dt$ of new juniors is recruited (where $J\hat{\chi} = \chi^S + J\chi^J$). The flow of talented majority (minority) candidates is Xdt. We will assume that $X \leq J\hat{\chi}$ (otherwise the organization would be homogenous, and the absence of minority juniors would deprive us of an analysis of the glass ceiling). Seniors have control over hiring and promotion decisions.

A glass ceiling in such hierarchical organizations results from control being located at the senior level. This operates through two channels:

- *Concern for control:* as earlier in the paper, control allows groups to engage in favoritism. Because control is located at the senior level, this in turn implies some discrimination in promotions, which in general exceeds that at the hiring level (if any). As noted in Section 4.1, a concern for control and the concomitant discrimination may arise even in large organizations, either because of shocks, or because the talent pool is larger in the minority (see footnote 38).

[53] Our results would still hold if we assumed instead that "stored" candidates stopped re-applying following some Poisson process.

- *Differential mingling effect:* for organizational reasons, senior members tend to hang around more with senior members than with junior ones. Their homophily concerns are therefore higher for promotions than for hiring decisions.

Because the second effect is at this stage of the paper newer, we illustrate it through a simple example, which can be much enriched in ways that we later discuss. Assume that senior members enjoy (expected lifetime) homophily benefits from in-group senior and junior members, which we denote respectively by b^S and b^J. The differential mingling effect is captured by $b^S > b^J$. A fraction $x \leq 1/2$ of new hires are in-group talented juniors, and similarly for the out-group ones: $xJ\hat{\chi}dt = Xdt$. Talent is observed prior to hiring. A talented member brings quality benefits to seniors equal to s^J when junior, and $s^S > s^J$ when senior. Assume that $s^l > b^l$ at both levels $l \in \{J, S\}$, and that $s^S - s^J > b^S - b^J$ (these two conditions generalize the previous assumption that quality matters to the majority).

In this framework, majority members are never worried about losing control, as the promotion of those who wil bring them the highest net benefits will always be tilted in favor of in-group juniors. This leads us to focus on the *majority's pecking order*: A promotion yields discounted net benefit to a majority senior member equal to 1) $s^S - s^J + b^S - b^J$ in the case of an in-group talented member; 2) $s^S - s^J$ for an out-group talented member; 3) $b^S - b^J$ for an in-group untalented member; 4) 0 for an out-group untalented member. This pecking order implies that promotion decisions will be tilted in favor of in-group members (except in the non-generic case in which all talented juniors are promoted and no untalented one is). In contrast, the junior population is balanced in composition; indeed, there is no rationale for the majority to discriminate at the hiring state as long as $s^J > b^J$.

When $X < \chi^S < 2X$, i.e. equivalently $x < 1/\left[1 + J\chi^J/\chi^S\right] < 2x$, in steady state the organization promotes all talented in-group juniors, a fraction z of talented out-group juniors, and no untalented juniors. The flows in and out of the junior and senior pools must balance, yielding respectively: $J\hat{\chi} = \chi^S + J\chi^J$, and $J\hat{\chi}x(1+z) = \chi^S$.

We define the glass ceiling index as the relative probability of promotion of talented majority and minority members, minus 1:[54]

$$\gamma \equiv \frac{1}{z} - 1 = \frac{2X - \chi^S}{\chi^S - X} \in (0, \infty)$$

In this region, the glass ceiling index is invariant with how pyramidal the organization is (J)[55], decreases with the frequency of senior-level vacancies (χ^S) and increases with the flow of talented candidates (X). Covering all parameter regions, the glass ceiling index is monotonic with χ^S/X.[56]

Proposition 10. ***(Glass ceiling)*** *In the hierarchical organization's steady state, hiring at the junior level is meritocratic*[57]*. By contrast, there exists a glass ceiling for minority juniors.*

[54] This definition of the glass ceiling index only looks at flows and is a conservative estimate of the glass ceiling; indeed, were we to look at stock, the glass ceiling effect would be stronger because the share of talented minority juniors promoted to seniority (over the whole stock of such juniors) would be below z (whenever $z < x$, the steady state of the junior pool features a mixture of talented minority and untalented majority juniors).

[55] An increase in J has two opposite effects: it makes it more difficult for a junior to be promoted, and talented minority members are the first to be left out; but it also makes talented juniors scarcer in the junior pool, increasing the minority members' probability of promotion.

[56] Indeed, for $\chi^S > 2X$, the senior majority hires all talented juniors and (some) untalented in-group juniors, and thus $\gamma = 0$, whereas for $\chi^S < X$, it promotes no out-group talented juniors, only talented in-group ones, and thus we set $\gamma = +\infty$.

[57] In line with Carmichael (1988) and Friebel-Raith (2004), it is thus optimal for the seniors' majority not to let current

This environment can be enriched in interesting ways. First, one may distinguish between talent and "senior potential"; only a fraction of talented members have the potential to make a more important contribution at the senior level; furthermore it may take time for the organization to discover who has such senior potential (there is a time of reckoning). Second, talented members may have outside opportunities, as in Section 4. Talented women may then quit the organization due to a discouragment effect: either they have been identified as lacking senior potential (their male counterparts by contrast staying in the organization), or the delay in being promoted is not worth the wait. Finally, the possibility of outside recruitment at the senior level would impact the glass ceiling effect.

6 Policy

We investigate the consequences of different interventions a principal could carry on. We successively consider two distinct informational environments:

(i) The principal at least occasionnally observes the candidates' talent but does not observe the candidates' and members' horizontal types. The principal may indeed have more information about quality than about horizontal attributes: a provost may use external letters or a visiting committee to assess the quality of a department or candidates, and a government may use a research assessment exercise to evaluate a university or its components. By contrast the provost, say, may not know whether the department is hiring buddies or researchers in a declining field that is their own. In a number of countries (such as France) the hiring of civil servants is merit based, in the sense that new civil servants take a competitive exam; in the US by contrast, "civil servants" are often political appointees. Exams for entering the civil service can be viewed as obtaining a signal of s that is informative.

(ii) The principal observes the candidates' and members' horizontal types, but only the incumbent members observe talent. Many public interventions such as affirmative action policies are based on gender, race, disability or religion, but not necessarily on quality.

In what follows, we take efficiency (quality plus homophily benefits), as the principal's objective function. Namely, denoting by S the organization's aggregate quality, by B the aggregate homophily benefits, and by T the transfer (if any) to the organization with ξ the cost of public funds, the principal's objective writes

$$W \equiv qS + B - \xi T$$

where $q \geq 1$ is the (relative) weight on quality.

In order to resolve the multiplicity of equilibria when two equilibria coexist in the absence of intervention (cf. Proposition 1), we assume that members coordinate on the Pareto-dominating meritocratic equilibrium (cf. Proposition 2).

Anticipating on the formal analysis, the consequences of the interventions we investigate are guided by three main forces:

- *Control is less appealing*: the constraints put on the majority's freedom reduce the value of control, favoring meritocracy over entrenchment.

juniors coopt new juniors as a majority of out-group juniors may engage in un-meritocratic hiring in order to increase their chances of being appointed to the senior board. This optimality result may not hold if for instance, juniors are better able than seniors at scouting talented candidates.

- *Quality matters more* (in the case of talent-based interventions), which also favors meritocracy.
- *Entrenched majorities may fight back*: on the other hand, if the intervention has milder consequences for larger majorities, then the intervention may promote super-entrenchment.

6.1 Quality-based interventions

Assuming that the principal has only talent information, we consider two policies by the principal: *(i)* stepping in by choosing the new member (which does not assume commitment by the principal), and *(ii)* rewarding quality (which does as the intervention is backward looking).

Discretionary overruling of majority decisions. We assume that in each period, the majority selects among the two candidates and then the principal can overrule the majority and pick the losing candidate. None of the players (principal, majority, and minority) can commit.

Suppose that the principal occasionally receives a signal. Namely, in each period the principal learns the quality of the candidates (or at least their quality differential) with probability λ and receives no signal with probability $(1-\lambda)$.

We look for equilibria (a) with level of entrenchment $l \geq 0$, and (b) in which the principal overrules the majority if and only if informed that the majority is violating meritocracy. Hence the probability η of intervention is given by $\eta = \lambda x$ if $M \leq k + l$, and $\eta = 0$ otherwise. The equilibrium nature of this intervention policy is motivated by the observations that (i) it is indeed efficient for the principal not to intervene without information, and (ii) the principal can obtain a one-period benefit when informed that meritocracy is violated.

For talent-blind discretionary interventions ($\lambda = 0$), it is an equilibrium for the principal not to engage in quality-blind interventions[58]; and so the meritocratic and entrenchment equilibria exist for the same parameter values as in the absence of intervention. So the impact of external interventions is here tied to the availability of evaluative information. When $\lambda = 1$, the principal can select the best candidate in each period, and there is no real "cooptation". So let us assume that $0 < \lambda < 1$.

The condition of existence of a meritocratic equilibrium is unchanged, as the principal has no reason to intervene in such an equilibrium. This property however does not hold for the entrenchment equilibrium. Intuitively, the possibility of intervention has two opposite effects on the principal's welfare. In the absence of behavioral response by the majority, the principal can occasionally overrule the majority and impose the meritocratic choice. But the majority may become wary of losing control when $M = k$ and so may decide to be super-entrenched so as to lower the probability of its losing control (without annihilating it completely, which is impossible).

The next proposition establishes that "well-meaning policies" systematically backfire for s/b close to 1 by generating full entrenchment.

Proposition 11. ***(Perverse effects of discretionary quality-based interventions)*** *Let $x < 1/2$.*

[58] The intuition for the result stems from two observations: (a) from the perspective of the principal (with $q \geq 1$), the majority takes the socially optimal decision for any majority size $M \geq k+1$, and if it is meritocratic, also when $M = k$, whereas if it is entrenched and tight, it takes the optimal decision with probability $1 - x \geq 1/2$; (b) if the majority is entrenched and tight, then its choice of candidate reveals no information on the latter's quality to the principal, and thus a talent-blind principal cannot outperform the majority's choice.

(i) The possibility of an informed overruling of majority decisions (with a strictly positive probability) results in full entrenchment for any s/b in a non-empty neighbourhood of 1, with $s/b \geq 1$.

(ii) The principal might achieve a higher welfare if it could commit not to intervene.

Remark. Fixing s/b, the principal's (ergodic aggregate) payoff increases with its probability λ of being informed on any interval such that the Pareto-dominating equilibrium remains unchanged. As the latter changes (a higher λ generating a higher level of entrenchment), the principal's payoff drops to a strictly lower value (see Figure 5 in Appendix M).

Rewarding quality. We now assume that the principal implements a quality assessment exercise according to a Poisson process of rate η, and this after the period-t election. A quality assessment exercise in period t results in an end-of-period bonus accruing to the organization and shared equally among the N members. We assume the bonus is one-shot, i.e. it is received at date t. Alternatively we could have assumed the bonus is split across several periods, yet frontloading the bonus is more effective[59]. For the sake of simplicity, we also assume the bonus is linear in the number of talented members in the organization: for each talented member in the organization at the end of period t, each member receives y. Consequently, the expected incremental lifetime contribution of a new talented (relative to mediocre) addition to each current member of the organization now writes as

$$s^+(\eta, y) \equiv s + \eta \frac{y}{1 - \delta_0(1 - 2/N)} = s\left(1 + \eta \frac{y}{\tilde{s}}\right) > s$$

while the expected lifetime utility for an incumbent member generated by the homophily payoff per new member sharing their opinion is still given by b.[60]

Proposition 12. ***(Rewarding quality)*** *For any positive cost of public funds ξ , there exists $\rho_\xi \in [1, \rho^m)$, strictly increasing with ξ and satisfying $\rho_0 = 1$, such that quality assessment exercises raise welfare (measured by ergodic quality minus cost of public funds, i.e. with per-period welfare $W = qS + B - \xi T$) if and only if $s/b \in [\rho_\xi, \rho^m)$.*

The intuition behind Proposition 12 is that for high s/b, the organization embraces meritocracy by itself and so spending public funds is wasteful. When instead the organization has little appetance for meritocracy (s/b small), the principal must pour large amounts of money on the organization to be effective, and this may prove too costly. It is thus only in the intermediate range that a boost promotes meritocracy and quality at a reasonable cost.[61]

6.2 Affirmative action

Suppose that the principal mandates diversity by setting a "representation threshold" – i.e. committing to imposing that the minority count at least R members at the end of any given period. Since it

[59] Because members may quit – and thus $\delta \leq (N-1)/N < 1$ –, frontloading the bonus maximizes the incentive for good recruitment.

[60] Computations go through as in the main model with a quality-payoff-over-homophily-benefit ratio now given by s^+/b instead of s/b. Hence, whenever $\rho^e < \infty$, for η, y sufficiently high, the ratio s^+/b is sufficiently high for the organization to reach the region where the unique symmetric MPE in weakly undominated strategies is the meritocratic equilibrium.

[61] We conjecture that when organizations compete with each other for talent (as in Section 4.2), rewarding quality may destabilize competition and create a mediocrity trap for the weakest and less diversified ones.

is suboptimal for the principal to impose parity[62], we focus on weaker forms of affirmative action with representation thresholds $R \leq k-1$.

Quality is reduced if at the moment of the vote, the representation threshold binds (i.e. $M = N - R$) and the majority candidate is more talented. Moreover, homophily benefits are also reduced on average. However there is an indirect effect: control is less appealing both because the majority is constrained and because the minority is favored. That effect might make the "constrained meritocratic" equilibrium (the choice is meritocratic except perhaps when $M = N - R$ at the moment of the vote) more likely, which might actually benefit the principal.

However, when s/b is very high, the efficiency loss at $M = N - R$ becomes extremely costly and majority members might be willing to pick the minority candidate at lower majority sizes whenever the latter is as talented as the majority one in order to avoid reaching a majority size of $M = N - R$ at a later period. We refer to such an equilibrium as *meritocracy with reverse favoritism*: members vote for their candidate if and only if he or she is strictly more talented than the rival candidate. In other words, the reverse-favoritism meritocratic (resp. constrained meritocratic) equilibrium features the most talented candidate being recruited with ties broken in favour of the minority (resp. majority) candidate – motivating its name[63]. Lastly, note that the (constrained) reverse-favoritism meritocracy equilibrium and the (constrained) meritocratic one are equivalent in terms of current-period efficiency for a given majority size, yet not in terms of average efficiency as the induced paths over majority sizes differ: reverse-favoritism meritocracy is on average more efficient.

Yet, affirmative action comes at a cost, both in terms of efficiency and homophily. Assuming $s/b < \rho^{\mathrm{m}}$, we compare the ergodic aggregate welfare in the entrenchment equilibrium under *laissez-faire* and the meritocratic equilibrium under affirmative action of level R. So entrenchment is the unique equilibrium under laissez-faire, while constrained meritocracy is also an equilibrium under affirmative action, further assuming meritocracy is selected whenever it co-exists with entrenchment in equilibrium.[64]

Proposition 13. ***(Affirmative action)***

(i) Existence regions. *Affirmative action in the form of a representation threshold $R \leq k-1$ expands the existence region of meritocracy.*[65] *Furthermore, for s/b sufficiently high, the unique meritocratic equilibrium*[66] *is meritocracy with reverse favoritism: the majority always selects the most talented candidate and breaks ties in favour of the minority candidate.*

[62] Suppose that the principal imposes parity (so at the end of the period the two groups are equally represented). Then the average quality of the coopted member ($\overline{x}s$) is smaller than in both the entrenched and meritocratic equilibria and homophily benefits are minimized.

[63] In order to alleviate the labels, we may omit the epithete "constrained" when referring to these equilibria whenever there is no ambiguity.

[64] When $N = 4$, explicit computations yield that for the organization to be meritocratic with reverse favoritism under affirmative action, it would have to be meritocratic under laissez-faire, in which case affirmative action is strictly dominated by laissez-faire whenever the principal internalizes the homophily payoffs. This motivates our comparing constrained meritocracy (and not reverse-favoritism meritocracy) under affirmative action to entrenchment under laissez-faire.

[65] Whenever a representation threshold is implemented, we refer to the "existence region of meritocracy" as the set of values of s/b for which there exists an equilibrium in which meritocratic recruitments take place whenever possible. [For $N = 4$, the existence regions of the constrained meritocratic and reverse-favoritism meritocratic equilibria have a non-empty intersection. Yet the existence region of meritocracy may differ in general from the union of the existence regions of the constrained meritocratic and the reverse-favoritism meritocratic equilibria, as the two may not overlap. Nonetheless, we show that for intermediate values of s/b, there exist other "meritocratic" equilibria depending on how ties are broken – in particular, "mixed-favoritism" meritocratic equilibria where ties are broken in favour of the majority (resp. minority) candidate when the majority size is far from (resp. close to) its upper bound. Hence we show that the existence region of meritocracy as defined is a convex set, namely a half-line.]

[66] In the sense that the most talented candidate is always recruited, unless maybe when the representation threshold is reached.

(ii) Ergodic aggregate welfare. *(a) The homophily (ergodic aggregate) payoff is strictly lower in the meritocratic equilibrium under affirmative action with representation threshold R than in the entrenchment equilibrium under laissez-faire. (b) There exists $x_{AA}(R) \in (0, 1/2)$ such that for any $x \in \big(0, x_{AA}(R)\big)$ (resp. $x \in \big(x_{AA}(R), 1/2\big)$), the quality (ergodic aggregate) payoff is strictly lower (resp. strictly higher) in the meritocratic equilibrium under affirmative action with representation threshold R than in the entrenchment equilibrium under laissez-faire (the two being equal for $x = x_{AA}(R)$). The cutoff $x_{AA}(R)$ strictly increases with R: the higher the representation threshold, the thinner the range of correlations for which meritocracy under affirmative action dominates entrenchment under laissez-faire.*

Remark. Affirmative action policies may have further positive welfare effects if candidacies are endogenous (and if for instance, candidates choose whether to invest in their talents).[67]

6.3 Supermajority electoral rules

We last consider a set of policies which could be implemented by an uninformed principal, namely voting rules requiring at least $k + l$ votes for a candidate to be elected for a given $l \geq 1$. We refer to Appendix P for details and only state here our main results.

We assume that the principal does not observe the candidates' talent. In line with our assumption of an uninformed principal, we posit that if no candidate reaches the election threshold, the principal picks one among the two at random[68]. Consequently, the principal's blindess makes failing to reach the election threshold costly for majority members. Consistently with this section's previous insights[69], we show that for $x < 1/2$, (i) for s/b sufficiently close to 1, super-entrenchment at level l is the unique symmetric MPE in weakly undominated strategies such that a stronger majority makes (weakly) more meritocratic recruitments; (ii) for δ sufficiently low, the existence region of meritocracy widens with respect to *laissez-faire*.

7 Alleys for future research

The introduction covered the main insights of our analysis. Consequently, these concluding remarks will focus on some of the (many) areas that would benefit from future research.

(a) More than two groups and coalitions. While a two-group structure is natural in a number of environments, exercising control over appointments may require building up a majoritarian coalition in others. As is well-known from academic departments or politics, such coalitions may be unstable over time, as a partner in a coalition may be evicted for the benefit of another or may be wary that the dominant coalition group becomes hegemonic. Studying such dynamics may involve a quantum leap in the complexity of the analysis, but would be very rewarding.[70]

(b) Human capital investment. Section 4 showed how quickly an entrenched organization can disin-

[67] Moreover, whenever such a policy disentrenches the organization, the higher expected homophily benefit for minority candidates may more than compensate the potential loss in aggregate quality, and thus enable the organization to attract talented minority candidates at higher outside options than an entrenched organization under laissez-faire.

[68] Admittedly, the intervention may rather be not to appoint any candidate in such circumstances. The organization would then face a deadlock as groups engage in a war of attrition.

[69] Here, two opposite effects drive the results: (a) the principal's blind intervention if the supermajority is not reached may make meritocracy relatively more attractive and prevent the organization from being entrenched; (b) super-entrenchment at level l shields the entrenched majority from the principal's intervention.

[70] One could use the Shapley value in order to compute a group's ability to select a candidate.

tegrate when talented candidates have outside options. In the same spirit, one could enrich the model by adding an ex-ante investment in human capital, which increases the probability of being "talented". The new feature relatively to the endogenous-candidacy section would be that the availability of talented minorities (determined by their incentive to invest in human capital) would be a public good from the point of view of organizations, as it is reasonable to assume that agents do not invest in human capital with a specific organization in mind. There might be self-fulfilling outcomes in which minorities do not invest because organizations are homogenous, and organizations are homogenous because they cannot find competent minorities. The multiplicity would then differ from the traditional one encountered in the theory of statistical discrimination[71], which is based on asymmetric information about the talent of a prospective candidate.

(c) Quits and competition. Another variant on the theme of equilibrium-determined membership decisions on the agent side would allow members of the organization to depart from the organization they are a member of, either because the quality has fallen or because the organization has become more entrenched.

(d) Heterogeneous tastes for homophily. We assumed that members have similar preferences for homophily. This need not be the case. A specialization may then arise, in which agents sort themselves out in their applications between highly entrenched organizations (the gentlemen's clubs of England and the Commonwealth countries) and more tolerant/open structures.

(e) Heterogeneous time discount factors and internal structure of power. Members' heterogeneous horizons in the organization affect their willingness to invest for the future. As we showed, discrimination against (resp. affirmative action in favor of) the minority is an investment benefitting patient majority members when the organization is attractive (resp. experiences difficulty in attracting talented minority candidates). Would "older" members (i.e. with a shorter time horizon) be more meritocratic than "younger" members? Or, to the contrary, would the young be more willing than the old to engage in voluntary affirmative action as candidates have outside options? The internal structure of power may thus balance the old's stronger preference for meritocracy when control may switch, with the young's greater propensity to invest in affirmative action when the organization fails to be attractive to minority talented candidates.

(f) Integrity of quality assessment exercises. One of our insights on the policy side is that quality assessment exercises promote meritocracy and diversity, and that, leaving their cost aside, they do not generate the perverse entrenchment effects that plague some other interventions. We however presumed that these assessments were accurate. Casual empiricism suggest that integrity is not to be taken for granted. Dominant groups may control not only the organizations themselves, but also the panels that are supposed to assess them. At the same time, minority groups may be minorities not because they suffer from some innate trait that is unrelated to quality (gender, ethnicity...), but because they are perceived as lower-quality agents by the majority group. Mandating diversity in the assessment panels may then be less appealing than when horizontal traits are really perceived to be horizontal. Capturing this may require a diversity of beliefs as to what constitutes high-quality work, and would for example shed much light on how science progresses.

(g) Cooptation as manipulation. This paper assumes that individuals outside the organization have no capability of causing harm to it. Introducing the possibility that coopting outsiders may change their

[71] See e.g., Arrow (1973), Coate-Loury (1993), Loury (2002), Phelps (1972) and Rosen (1997).

behaviour and reduce nuisance would allow us to capture the second meaning of "cooptation" originating with Selznick.[72]

(g) Searching for talented candidates. Talented candidates, even if they are willing to join the organization, may not become members because they are unaware of an opening or have misconceptions about their chance of being coopted. Search raises a host of interesting questions: does it result from members' initiative or is it conducted through a search committee? In the former case, does the majority benefit from its larger size (which is unclear: a larger membership size increases the number of coincidental thoughts as well as the extent of free riding; social pressure within groups may also differ)? Is the intuition that search will be mainly directed toward in-group candidates correct?[73]

(h) Empirical investigations. Last, but certainly not least, the model could be tested, from its basic assumptions to its predictions. For instance, the homophily incentive b has in recent years increased in some dimensions (political polarization) and decreased in others (as when the law or social norms penalize a lack of diversity); depending on factors such as initial conditions, the nature of inside interactions, or the competitiveness of the talent market, this evolution should impact dependent variables such as the quality of recruitments, the heterogeneity among organizations and their divergent paths.

References

Acemoglu, D., Egorov, G., and Sonin, K. (2009). "Equilibrium Refinement in Dynamic Voting Games". mimeo.

Acemoglu, D., Egorov, G., and Sonin, K. (2012). "Dynamics and Stability of Constitutions, Coalitions, and Clubs". *American Economic Review*, 102(4):1446–1476.

Acemoglu, D. and Robinson, J. A. (2000). "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective". *The Quarterly Journal of Economics*, 115(4):1167–1199.

Arrow, K. (1973). "The Theory of Discrimination". In Ashenfelter, O. and Rees, A., editors, *Discrimination in Labor Markets*. Princeton University Press.

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

Bagues, M. and Esteve-Volart, B. (2010). "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment". *The Review of Economic Studies*, 77(4):1301–1328.

[72] One of many ways of capturing the cooptation of members with sufficient nuisance power outside is to assume that the probability that the organization continues falls sharply when it is too monolithic – e.g. due to the prospect of a "revolution".

[73] Suppose, say, that search is the prerogative of a committee and focuses on talented candidates (there are always untalented candidates of both groups). With probability $\overline{x}$, there is a talented candidate of a given group. Only the majority can search for candidates (one could also introduce search by the minority, although the latter might be less motivated because it only has real authority, no formal one). With probability e^+, search cost $\psi(e^+)$, the majority finds the talented majority candidate if there is one (so the overall probability of there being a talented majority candidate is $e^+\overline{x}$); and similarly, the probability of finding the talented minority candidate if there is one is e^-. The function ψ is the same for both searches, is increasing and convex, and it satisfies $\psi'(0) = 0$ and $\psi'(1) = +\infty$. Note that for costless search, we are back to our model. We conjecture that $e^+(M) > e^-(M)$ (with $e^-(k) = 0$ in an entrenched equilibrium) as (i) finding a talented minority candidate when a talented majority candidate has been identified is useless for the majority; and (ii) finding a talented majority candidate yields an extra b.

Bagues, M., Sylos-Labini, M., and Zinovyeva, N. (2017). "Does the Gender Composition of Scientific Committees Matter?". *American Economic Review*, 107(4):1207–38.

Barberà, S., Maschler, M., and Shalev, J. (2001). "Voting for Voters: A Model of Electoral Evolution". *Games and Economic Behavior*, 37(1):40–78.

Becker, G. (1957). *The Economics of Discrimination*. University of Chicago Press.

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

Bertocchi, G. and Spagat, M. (2001). "The Politics of Co-optation". *Journal of Comparative Economics*, 29(4):591–607.

Bertrand, M. (2018). "Coase Lecture – The Glass Ceiling". *Economica*, 85(338):205–231.

Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2018). "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway". *The Review of Economic Studies*, 86(1):191–239.

Board, S., Meyer-ter-Vehn, M., and Sadzik, T. (2019). "Recruiting Talent". mimeo, UCLA.

Buchanan, J. M. (1965). "An Economic Theory of Clubs". *Economica*, 32(125):1–14.

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.

Carmichael, L. (1988). "Incentives in Academics: Why is There Tenure?". *Journal of Political Economy*, 96(3):453–472.

Coate, S. and Loury, G. (1993). "Will Affirmative-Action Policies Eliminate Negative Stereotypes?". *American Economic Review*, 83(5):1220–1240.

Daley, D. J. (1968). "Stochastically Monotone Markov Chains". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 10(4):305–317.

Egorov, G. and Polborn, M. (2011). "An Informational Theory of Homophily". mimeo.

Friebel, G. and Raith, M. (2004). "Abuse of Authority and Hierarchical Communication". *RAND Journal of Economics*, 35(2):224–244.

Fryer, R. G. J. and Loury, G. C. (2005). "Affirmative Action and Its Mythology". *Journal of Economic Perspectives*, 19(3):147–162.

Hoffman, M., Kahn, L. B., and Li, D. (2017). "Discretion in Hiring". *The Quarterly Journal of Economics*, 133(2):765–800.

Jehiel, P. and Scotchmer, S. (1997). "Free Mobility and the Optimal Number of Jurisdictions". *Annals of Economics and Statistics*, (45):219–231.

Keilson, J. and Kester, A. (1977). "Monotone Matrices and Monotone Markov Processes". *Stochastic Processes and their Applications*, 5(3):231–241.

Loury, G. (2002). *The Anatomy of Racial Inequality*. Harvard University Press.

Maskin, E. and Tirole, J. (2001). "Markov Perfect Equilibrium, I: Observable Actions". *Journal of Economic Theory*, 100:191–219.

Mattozzi, A. and Merlo, A. (2015). "Mediocracy". *Journal of Public Economics*, 130:32–44.

Moldovanu, B. and Shi, X. (2013). "Specialization and Partisanship in Committee Search". *Theoretical Economics*, 8:751–774.

Phelps, E. S. (1972). "The Statistical Theory of Racism and Sexism". *American Economic Review*, 62(4):659–661.

Rivera, L. A. (2012). "Hiring as Cultural Matching: The Case of Elite Professional Service Firms". *American Sociological Review*, 77(6):999–1022.

Roberts, K. (2015). "Dynamic Voting in Clubs". *Research in Economics*, 69(3):320–335.

Rosén, A. (1997). "An Equilibrium Search-Matching Model of Discrimination". *European Economic Review*, 41(8):1589–1613.

Schmeiser, S. (2012). "Corporate board dynamics: Directors voting for directors". *Journal of Economic Behavior & Organization*, 82(2):505 – 524.

Selznick, P. (1948). "Foundations of the Theory of Organization". *American Sociological Review*, 13(1):25–35.

Selznick, P. (1949). *TVA and the Grass Roots: A Study in the Sociology of Formal Organization*. Berkely: University of California Press.

Sobel, J. (2000). "A Model of Declining Standards". *International Economic Review*, 41(2):295–303.

Tiebout, C. M. (1956). "A Pure Theory of Local Expenditures". *Journal of Political Economy*, 64:416–424.

Tirole, J. (1996). "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)". *The Review of Economic Studies*, 63(1):1–22.

Zinovyeva, N. and Bagues, M. (2015). "The Role of Connections in Academic Promotions". *American Economic Journal: Applied Economics*, 7(2):264–92.

Appendix

A Value functions for majority and minority members

Value function for a majority member. Under restrictions (i)-(ii) on the majority's strategy, let V_i^- denote the expected value function conditional on the minority candidate being more talented, and V_i^+ denote the expected value function conditional on the complementary event. The value function for a majority member (see below for that of a minority member) writes for any $k \leq M \leq N-1$[74],

$$V_M = xV_M^- + (1-x)V_M^+ \tag{8}$$

$$\text{where} \quad \begin{cases} V_M^- = \max\left\{ b + \delta\left(\frac{M}{N-1}V_M + \left(1 - \frac{M}{N-1}\right)V_{M+1}\right), \right. \\ \qquad\qquad \left. s + \delta\left(\frac{M-1}{N-1}V_{M-1} + \left(1 - \frac{M-1}{N-1}\right)V_M\right)\right\} \\ V_M^+ = b + \frac{\overline{x}}{1-x}s + \delta\left(\frac{M}{N-1}V_M + \left(1 - \frac{M}{N-1}\right)V_{M+1}\right) \end{cases}$$

With probability x, the majority faces a trade-off between selecting a talented minority member (yielding payoff s) and picking the less talented majority candidate (yielding payoff b). With probability $1-x$, the choice is a no-brainer and the majority candidate brings average payoff $b + \overline{x}s/(1-x)$ where $\overline{x}/(1-x)$ is the conditional probability of that candidate's being talented. Furthermore, for a majority member, recruiting a majority candidate when the majority has size M in period t yields an *ex-ante* (i.e. before departure) majority size of $M+1$. Three events might then happen at the beginning of period $t+1$ before the vote takes place: *(i)* with probability $1/N$ (which is already embodied in the discount factor δ), the majority member quits the organization, which gives him zero payoff; *(ii)* with probability M/N, *another* majority member quits, and thus the majority size decreases to M; *(iii)* with probability $(N-M-1)/N$, a minority member quits, and thus the majority size remains equal to $M+1$.

Value function for a minority member. If the majority recruits the majority candidate in period t, then at the beginning of period $t+1$: *(i)* with probability $1/N$, the minority member quits the organization, which gives her zero payoff; *(ii)* with probability $(M+1)/N$, a majority member quits, and thus the majority size decreases to M; *(iii)* with probability $(N-M-2)/N$, *another* minority member quits, and thus the majority size remains equal to $M+1$.

Let $\sigma(M) \in \{0,1\}$ be defined for any $k \leq M \leq N-1$ as

$$\sigma(M) = \begin{cases} 1 & \text{if } b + \delta\left(\frac{M}{N-1}V_M + \left(1 - \frac{M}{N-1}\right)V_{M+1}\right) > s + \delta\left(\frac{M-1}{N-1}V_{M-1} + \left(1 - \frac{M-1}{N-1}\right)V_M\right) \\ 0 & \text{otherwise.} \end{cases}$$

[74] Equation (8) applies even when $M = N-1$ as the majority size $M+1$ becomes irrelevant (its probability being nil).

The value function for a minority member writes for any $k \leq M \leq N-2$:

$$V_{N-M-1} = xV^-_{N-M-1} + (1-x)V^+_{N-M-1} \tag{9}$$

$$\text{where} \quad \begin{cases} V^-_{N-M-1} = \sigma(M)\delta\Big(\frac{M+1}{N-1}V_{N-M-1} + \big(1-\frac{M+1}{N-1}\big)V_{N-M-2}\Big) \\ \qquad\qquad +(1-\sigma(M))\Big[s+b+\delta\Big(\frac{M}{N-1}V_{N-M} + \big(1-\frac{M}{N-1}\big)V_{N-M-1}\Big)\Big] \\ V^+_{N-M-1} = \frac{\overline{x}}{1-x}s + \delta\Big(\frac{M+1}{N-1}V_{N-M-1} + \Big(1-\frac{M+1}{N-1}\Big)V_{N-M-2}\Big) \end{cases}$$

B Proof of Lemma 1

The result for $N=4$ derives from straightforward computations[75]. We assume in the following $N \geq 6$. Assume $s > b > 0$ and $x \in (0, 1/2)$.

Proof of (i). Consider, first, the entrenched equilibrium. We omit the e superscript in order to alleviate the notation. For any $M \in \{k-1, ..., N-2\}$, let $u_M \equiv V_{M+1} - V_M$. By writing the expression of the value function from (8) in $M \in \{k+1, ..., N-1\}$ (thus writing V_M as a function of V_{M-1}, V_M and V_{M+1}), and then substracting the expression in M from the expression in $M+1$ yields for any $M \in \{k+1, ..., N-2\}$[76],

$$\begin{aligned} V_{M+1} - V_M = \delta x\Big[\frac{M-1}{N-1}(V_M - V_{M-1}) + \Big(1-\frac{M}{N-1}\Big)(V_{M+1} - V_M)\Big] \\ + \delta(1-x)\Big[\frac{M}{N-1}(V_{M+1} - V_M) + \Big(1-\frac{M+1}{N-1}\Big)(V_{M+2} - V_{M+1})\Big] \end{aligned}$$

i.e. by rearranging the terms,

$$\Big[1 - \delta x\Big(1-\frac{M}{N-1}\Big) - \delta(1-x)\frac{M}{N-1}\Big]u_M = \delta x\frac{M-1}{N-1}u_{M-1} + \delta(1-x)\Big(1-\frac{M+1}{N-1}\Big)u_{M+1} \tag{10}$$

We show the result by contradiction. The intuition is as follows: if the majority were to prefer being in $N-2$ to $N-1$ (i.e. if $u_{N-2} \leq 0$), then by induction, it would prefer being in k over all majority sizes. Yet this is not possible as by definition of the entrenched equilibrium, the majority would then prefer to be in $k+1$ in order to avoid the current-period loss of efficiency due to entrenchment in k. Hence a contradiction and thus the majority prefers being in $N-1$ than in $N-2$. The result obtains with a similar induction argument yielding: $u_k > u_{k+1} > ... > u_{N-2} > 0$.

[75] Using (8) and (9), for the entrenched equilibrium, one has

$$\Big[1 - \frac{2\delta}{3}(1-x)\Big](V^e_3 - V^e_2) = x(s-b)$$

and thus $V^e_1 = \overline{x}s/(1-\delta) < (b+\overline{x}s)/(1-\delta) < V^e_2 < V^e_3$. Similarly for the meritocratic equilibrium:

$$\begin{cases} \Big[1 - \frac{x\delta}{3} - \frac{2\delta}{3}(1-x)\Big](V^m_3 - V^m_2) = \frac{x\delta}{3}(V^m_2 - V^m_1) \\ \Big[1 - \delta(1-x)\Big](V^m_2 - V^m_1) = (1-2x)b + \delta\frac{(1-x)}{3}(V^m_3 - V^m_2) \end{cases}$$

and thus $V^m_1 < V^m_2 < V^m_3$, and $V^m_2 - V^m_1 > V^m_3 - V^m_1$.

[76] With the abuse of notation $u_{N-1} = 0$, which is irrelevant since the coefficient of u_{N-1} is nil.

Suppose $u_{N-2} \leq 0$. Then Equation (10) for $M = N-2$ implies

$$\left[1 - \delta\frac{x}{N-1} - \delta(1-x)\frac{N-2}{N-1}\right]u_{N-2} = \delta x\frac{N-3}{N-1}u_{N-3}$$

Therefore, $u_{N-3} \leq 0$ and $u_{N-3} \leq u_{N-2}$. We then proceed by induction to show that for any $M \in \{k+1, ..., N-2\}$, $u_{M-1} \leq u_M \leq 0$. Assume the result holds for all indices in $\{M+1, ..., N-2\}$. Then Equation (10) implies

$$\left[1 - \delta x\left(1 - \frac{M}{N-1}\right) - \delta(1-x)\frac{M}{N-1}\right]u_M \geq \delta x\frac{M-1}{N-1}u_{M-1} + \delta(1-x)\left(1 - \frac{M+1}{N-1}\right)u_M$$

$$\text{i.e.} \quad \left[1 - \delta x\left(1 - \frac{M}{N-1}\right) - \delta(1-x)\frac{N-2}{N-1}\right]u_M \geq \delta x\frac{M-1}{N-1}u_{M-1}$$

Consequently, $u_{M-1} \leq u_M \leq 0$. Hence the result by induction. In particular, one has $u_k \leq 0$, i.e. $V_{k+1} - V_k \leq 0$. However, writing Equation (8) in $k+1$ and k and taking the difference yields

$$u_k = x(s-b) + \delta(1-x)\left[\left(1 - \frac{k+1}{N-1}\right)u_{k+1} + \frac{k}{N-1}u_k\right] \tag{11}$$

and thus

$$0 \geq \left[1 - \delta(1-x)\frac{N-2}{N-1}\right]u_k \geq x(s-b) > 0,$$

which is a contradiction. Therefore $V_{N-1} - V_{N-2} = u_{N-2} > 0$. It is then easy to see that the same induction argument used above, using repeatedly Equation (10), shows that $u_{M-1} > u_M > 0$ for any $M \in \{k+1, ..., N-2\}$. Hence the result for the entrenched equilibrium.

Consider now the meritocratic equilibrium. We again omit the superscript on the value function in order to alleviate the notation. Let $u_i \equiv V_{i+1} - V_i$ for any $i \in \{1, ..., N-2\}$. Note that Equation (10) holds for any $M \in \{k, ..., N-2\}$. The argument is similar to the one used in the entrenched equilibrium: the idea is again to suppose that the majority prefers being in $N-2$ over $N-1$ and reach a contradiction. Note that there are two differences with respect to the entrenchment setup, as *(a)* the contradiction stems from the loss of homophily payoff (when candidates have same talent) associated with losing the majority, and *(b)* since there is no entrenchment and majorities switch, in order to reach the contradiction, the induction needs to go down until a group size of 1.

Assume by contradiction that $u_{N-2} \leq 0$. Then, by induction, this implies that for any $M \in \{k, ..., N-2\}$, $u_{M-1} \leq u_M \leq 0$, and thus in particular $u_{k-1} \leq u_k \leq 0$.

Consider now u_1. Note that by writing the expression of the value function from (9) in $M \in \{k+1, ..., N-1\}$ (thus writing V_{N-M-1} as a function of V_{N-M-2}, V_{N-M-1} and V_{N-M}), and then substracting the expression in $(N-M-1)$ from the expression in $(N-M)$ (and rearranging) yields for any $M \in \{k+2, ..., N-2\}$:

$$\left[1 - \delta(1-x)\frac{M}{N-1} - \delta x\left(1 - \frac{M}{N-1}\right)\right]u_{N-M-1} = \delta(1-x)\left(1 - \frac{M+1}{N-1}\right)u_{N-M-2} + \delta x\frac{M-1}{N-1}u_{N-M} \tag{12}$$

and in particular,

$$\left[1-\delta\frac{x}{N-1}-\delta(1-x)\frac{N-2}{N-1}\right]u_1=\delta x\frac{N-3}{N-1}u_2$$

By the usual induction argument using (12), $u_1>0$ implies $0<u_1<u_2<...<u_{k-2}<u_{k-1}$, which contradicts $u_{k-1}\leq 0$. Hence $u_1\leq 0$ and the same induction argument now implies $0\geq u_1\geq u_2\geq ...\geq u_{k-2}\geq u_{k-1}$.

However, substracting Equation (8) in k and Equation (9) in $k-1$ yields after rearranging:

$$\left[1-\delta(1-x)\right]u_{k-1}=(1-2x)b+\delta(1-x)\left[\left(1-\frac{k}{N-1}\right)u_k+\left(1-\frac{k+1}{N-1}\right)u_{k-2}\right] \tag{13}$$

The contradiction then obtains by summing the above equation together with Equations (10) and (12) over all indexes (and rearranging), which gives:

$$\left(1-\delta\frac{x}{N-1}-\delta(1-x)\right)(u_1+u_{N-2})+\big(1-\delta\big)\sum_{i=2}^{N-3}u_i=(1-2x)b>0$$

This contradicts the fact that $u_i\leq 0$ for all $i\in\{1,...,N-2\}$. Therefore $u_{N-2}>0$. Using repeatedly the usual induction argument with Equation (10) yields the result.

The proof of claim *(ii)* relies on the same arguments as the proof of *(i)* and is thus omitted for the sake of brevity.

Claim *(iii)* derives from arguments analogous to the ones used in the proofs of *(i)* and *(ii)*. The result is obvious with *(i)* for the meritocratic equilibrium. The result for the entrenchment equilibrium obtains by considering the sequence V_i-V_{N-1-i} for $i\in\{k,...,N-2\}$ and using (8)-(9).[77]

Suppose by contradiction that $V_k-V_{k-1}<0$. This implies that $V_{k+1}-V_{k-2}<V_k-V_{k-1}<0$, and thus by induction that $V_{N-1}-V_1<V_{N-2}-V_1<...<V_k-V_{k-1}<0$, which contradicts $V_{N-1}\geq V_{N-2}$ as shown above. (Another contradiction would be reached by summing as above the analogues of (10)-(12) and noting that the RHS is positive whenever $x\leq 1/2$). Hence $V_k-V_{k-1}\geq 0$. If $V_{k+1}-V_{k-2}<0$, the same contradiction is reached again as then $V_{N-1}-V_1<V_{N-2}-V_1<...<V_{k+1}-V_k<0$ (Again, one could sum over $i\in\{k+1,...,N-2\}$ the analogues of (10)-(12) and note that the RHS is positive whenever $x\leq 1/2$). The result obtains by induction: for any $i\in\{k,...,N-2\}$, $V_i-V_{N-1-i}\geq 0$. Details of the proof show that the inequality is strict if and only if $b>0$, or $x>0$ and $s>0$.

[77] Namely, for $M\in\{k+1,...,N-3\}$,

$$\left[1-\delta(1-x)\frac{M}{N-1}-\delta x\Big(1-\frac{M-1}{N-1}\Big)\right](V_M-V_{N-M-1})-(1-2x)b+\frac{\delta}{N-1}\Big[(1-x)u_{N-M-2}+xu_{N-M-1}\Big]$$
$$=\delta(1-x)\Big(1-\frac{M}{N-1}\Big)(V_{M+1}-V_{N-M-2})+\delta x\frac{M-1}{N-1}(V_{M-1}-V_{N-M})$$

while for $M=k$ and $M=N-2$,

$$\left[1-\delta\frac{k}{N-1}\right](V_k-V_{k-1})=b-\frac{\delta}{N-1}u_{k-2}+\delta\Big(1-\frac{k}{N-1}\Big)(V_{k+1}-V_{k-2}),$$
$$\left[1-\delta(1-x)\frac{N-2}{N-1}-\delta x\frac{2}{N-1}\right](V_{N-2}-V_1)=(1-2x)b-\frac{\delta x}{N-1}u_1+\delta\frac{(1-x)}{N-1}(V_{N-1}-V_1)+\delta x\frac{N-3}{N-1}(V_{N-3}-V_2)$$

Recall that in the entrenched equilibrium, $u_i\leq 0$ for any $i\leq k-2$, strictly so if and only if $x(s+b)>0$.

C Proof of Proposition 1

C.1 Proof of Proposition 1-(i)

The proof unfolds in two steps. Firstly we show that for any V_{k-1} in a compact convex set, there exists a unique equilibrium sequence of majority value functions $(V_k, ..., V_{N-1})$. Secondly we show that, for any V_{k-1}, either the entrenched or the meritocratic strategies satisfy the Bellman equations. The two steps combined establish that for any V_{k-1}, only the entrenched or meritocratic strategies can satisfy the Bellman equations and that exactly one of them does so, yielding the result. Let v (resp. w) denote the value brought to a member of the majority by the minority (resp. majority) candidate. So $v \in \{0, s\}$ and $w \in \{b, b+s\}$.

Step 1. Consider the compact convex set $\mathcal{C} \equiv \big[0, (s+b)/(1-\delta)\big]^k$. All vectors of value function $(V_k, .., V_{N-1})$ necessarily belong to $\mathcal{C}$. Fix an arbitrary value of $V_{k-1} \in \mathcal{C}$ and consider the mapping T from $\mathcal{C}$ into itself, $T : V \equiv (V_k, ..., V_{N-1}) \longmapsto TV \equiv (TV_k, ..., TV_{N-1})$ defined for any $i \geq k$ by

$$\begin{aligned} TV_i &\equiv \mathbb{E}_{v,w}\left[\max\left\{v + \delta\left(\frac{i-1}{N-1}V_{i-1} + \left(1 - \frac{i-1}{N-1}\right)V_i\right), w + \delta\left(\frac{i}{N-1}V_i + \left(1 - \frac{i}{N-1}\right)V_{i+1}\right)\right\}\right] \\ &\equiv \mathbb{E}_{v,w}[\tilde{V}_i(v,w)] \end{aligned}$$

We show that T is a contraction mapping. We use the norm defined for any $V \in \mathbb{R}^k$ by $\|V\| \equiv \max_i |V_i|$. Let $V, V' \in \mathcal{C}$ and consider a given realization of (v, w) and $i \in \{k, ..., N-1\}$. Three different cases may arise:

- the inequality: $v + \delta\left[\frac{i-1}{N-1}V_{i-1} + \left(1 - \frac{i-1}{N-1}\right)V_i\right] \geq w + \delta\left[\frac{i}{N-1}V_i + \left(1 - \frac{i}{N-1}\right)V_{i+1}\right]$ holds for both sequences V and V', and thus

$$|\tilde{V}_i'(v,w) - \tilde{V}_i(v,w)| = \delta\left|\frac{i-1}{N-1}(V'_{i-1} - V_{i-1}) + \left(1 - \frac{i-1}{N-1}\right)(V'_i - V_i)\right| \leq \delta\|V' - V\|$$

- the inequality: $v + \delta\left[\frac{i-1}{N-1}V_{i-1} + \left(1 - \frac{i-1}{N-1}\right)V_i\right] \leq w + \delta\left[\frac{i}{N-1}V_i + \left(1 - \frac{i}{N-1}\right)V_{i+1}\right]$ holds for both sequences V and V', and thus

$$|\tilde{V}_i'(v,w) - \tilde{V}_i(v,w)| = \delta\left|\frac{i}{N-1}(V'_i - V_i) + \left(1 - \frac{i}{N-1}\right)(V'_{i+1} - V_{i+1})\right| \leq \delta\|V' - V\|$$

- otherwise, the following inequalities hold (possibly inverting the roles of V and V'):

$$\begin{cases} v + \delta\left[\frac{i-1}{N-1}V_{i-1} + \left(1 - \frac{i-1}{N-1}\right)V_i\right] \leq w + \delta\left[\frac{i}{N-1}V_i + \left(1 - \frac{i}{N-1}\right)V_{i+1}\right] \\ v + \delta\left[\frac{i-1}{N-1}V'_{i-1} + \left(1 - \frac{i-1}{N-1}\right)V'_i\right] \geq w + \delta\left[\frac{i}{N-1}V'_i + \left(1 - \frac{i}{N-1}\right)V'_{i+1}\right] \end{cases} \tag{14}$$

and thus

$$\begin{cases} \tilde{V}_i(v,w) = w + \delta\left[\dfrac{i}{N-1}V_i + \left(1 - \dfrac{i}{N-1}\right)V_{i+1}\right] \\ \tilde{V}'_i(v,w) = v + \delta\left[\dfrac{i-1}{N-1}V'_{i-1} + \left(1 - \dfrac{i-1}{N-1}\right)V'_i\right] \end{cases}$$

Then,

- either $\tilde{V}'_i(v,w) \geq \tilde{V}_i(v,w)$, and by the first inequality in (14),

$$|\tilde{V}'_i(v,w) - \tilde{V}_i(v,w)| \leq \delta\left|\frac{i-1}{N-1}(V'_{i-1} - V_{i-1}) + \left(1 - \frac{i-1}{N-1}\right)(V'_i - V_i)\right| \leq \delta\|V' - V\|,$$

- or $\tilde{V}'_i(v,w) < \tilde{V}_i(v,w)$, and by the second inequality in (14),

$$|\tilde{V}'_i(v,w) - \tilde{V}_i(v,w)| \leq \delta\left|\frac{i}{N-1}(V'_i - V_i) + \left(1 - \frac{i}{N-1}\right)(V'_{i+1} - V_{i+1})\right| \leq \delta\|V' - V\|$$

Therefore, taking the expectation over all realizations of (v,w) and taking the maximum over indices yields: $\|TV' - TV\| \leq \delta\|V' - V\|$ (where $\delta < 1$). Hence T is a contraction mapping over the compact convex set $\mathcal{C}$, and thus by Banach fixed-point theorem, it has a unique fixed point. Note that since T depends on V_{k-1}, we have shown that for any V_{k-1}, there exists a unique equilibrium sequence of majority value functions $V_k, ..., V_{N-1}$.

Step 2. We first prove the following lemma:

Lemma 3. *Conditional on there being no profitable deviation when $M = k$ and the minority candidate is strictly more talented, there exists no profitable one-shot deviation from a canonical strategy (i) at any majority size and whenever the majority candidate is at least as talented as the minority candidate, and (ii) at any majority size $M \geq k+1$ and whenever the minority candidate is strictly more talented.*

Proof. In both canonical equilibria, $u_i \equiv V_{i+1} - V_i$ is positive and decreasing for $i \geq k$ from Lemma 1. Hence, for any profile of candidates' "values" $(v,w) \in \{(0,b),(0,b+s),(s,b+s)\}$, this implies that for any $M \geq k$,

$$\delta\left(1 - \frac{M}{N-1}\right)\left[V_{M+1} - V_M\right] + \delta\frac{M-1}{N-1}\left[V_M - V_{M-1}\right] \geq 0 \geq v - w,$$

and thus by construction, the canonical strategies are optimal at any majority size whenever the majority candidate is at least as talented as the minority one: the majority then optimally selects its own candidate. Hence it remains to show that the canonical strategies are optimal at any majority size $M \geq k+1$ whenever the minority candidate is strictly more talented ($(v,w) = (s,b)$).

The proof unfolds in two steps:

(a) Entrenchment equilibrium: we show that, conditional on entrenchment at $M = k$ being optimal, the canonical entrenchment strategy is optimal at any other majority size and any other candidates' vertical types.

(b) Meritocratic equilibrium: we show similarly that meritocracy is optimal at any majority size conditional on being optimal at $M = k$.

(a) Assume entrenchment is optimal when $M = k$, i.e. the majority is better off voting for an untalented majority candidate against a talented minority one. Then, letting V^{e} denote the value function in the entrenchment equilibrium – hence with V^{e}_{k-1} determined by the minority's entrenchment strategy –, (8) implies:

$$\delta\left(1-\frac{k}{N-1}\right)\left[V^{\mathrm{e}}_{k+1}-V^{\mathrm{e}}_{k}\right]+\delta\frac{k-1}{N-1}\left[V^{\mathrm{e}}_{k}-V^{\mathrm{e}}_{k-1}\right]\geq s-b,$$

i.e. using the notation $u^{\mathrm{e}}_i \equiv V^{\mathrm{e}}_{i+1}-V^{\mathrm{e}}_i$,

$$\delta\frac{k-1}{N-1}u^{\mathrm{e}}_k+\delta\frac{k-1}{N-1}u^{\mathrm{e}}_{k-1}\geq s-b$$

Similarly, it is optimal for the majority to recruit a talented minority candidate against an untalented majority one at majority size M if and only if:

$$\delta\left(1-\frac{M}{N-1}\right)u^{\mathrm{e}}_M+\delta\frac{M-1}{N-1}u^{\mathrm{e}}_{M-1}\leq s-b,$$

Equation (11) implies that

$$\delta\left(\frac{k-2}{N-1}u^{\mathrm{e}}_{k+1}+\frac{k}{N-1}u^{\mathrm{e}}_k\right)=\frac{1}{1-x}\left(u^{\mathrm{e}}_k-x(s-b)\right)$$

Using the previous equation together with the inequality $u^{\mathrm{e}}_{k+1}\leq u^{\mathrm{e}}_k$ from Lemma 1 yields

$$\left[1-\delta(1-x)\frac{N-2}{N-1}\right]u^{\mathrm{e}}_k\leq x(s-b)$$

Therefore, since $\delta<(N-1)/N$,

$$\delta\left(\frac{k-2}{N-1}u^{\mathrm{e}}_{k+1}+\frac{k}{N-1}u^{\mathrm{e}}_k\right)\leq\frac{\delta x\frac{N-2}{N-1}}{1-\delta(1-x)\frac{N-2}{N-1}}(s-b)<s-b,$$

and hence it is indeed optimal for the majority to pick a talented minority candidate against an untalented majority one when $M=k+1$. The result extends to any majority size $M\geq k+1$ by monotonicity of the sequence $(u^{\mathrm{e}}_i)_i$ which decreases with respect to i.

(b) Assume meritocracy is optimal when $M = k$, i.e. the majority is better off voting for a talented minority candidate against an untalented majority one. Hence, with the usual notation, (8) implies:

$$\delta\frac{k-1}{N-1}u^{\mathrm{m}}_k+\delta\frac{k-1}{N-1}u^{\mathrm{m}}_{k-1}\leq s-b$$

Therefore, by Lemma 1, the monotonicity of the sequence $(u^{\mathrm{m}}_i)_i$ which decreases with respect to i yields that for any $M\geq k+1$,

$$\delta\left(1-\frac{M}{N-1}\right)u^{\mathrm{m}}_M+\delta\frac{M-1}{N-1}u^{\mathrm{m}}_{M-1}\leq s-b,$$

which concludes the proof of the Lemma. □

Fix V_{k-1} and consider the unique equilibrium sequence of majority value functions $V_k, ..., V_{N-1}$ solving the Bellman equations with V_{k-1}. Consider majority size k. For $s > b$, exactly one of the following equations holds:

$$\text{either} \qquad V_k = \mathbb{E}_w[w] + \delta\left[\frac{k}{N-1}V_k + \left(1 - \frac{k}{N-1}\right)V_{k+1}\right] \tag{15}$$

$$\text{or} \qquad V_k > \mathbb{E}_w[w] + \delta\left[\frac{k}{N-1}V_k + \left(1 - \frac{k}{N-1}\right)V_{k+1}\right] \tag{16}$$

Note that (15) (resp. (16)) is equivalent to entrenchment (resp. meritocracy) in k in the sense that an untalented majority candidate is elected (resp. not elected) against a talented minority one.

Assume (15) holds. Then Lemma 3 (see below) applies with entrenchment being optimal in $M = k$. Therefore the value function of the entrenched equilibrium solves the Bellman equations. By step 1 (since T is a contraction mapping), we further have that the entrenched strategy is the unique strategy that satisfies the Bellman equations for V_{k-1} such that (15) holds.

Conversely, a similar argument relying on Lemma 3 (and ultimately on Lemma 1) shows that, for V_{k-1} such that (16) holds, the meritocratic strategy is the unique strategy that satisfies the Bellman equations. In conclusion, for each value of V_{k-1}, either the entrenched or the meritocratic strategy solves the Bellman equations.

Therefore, all symmetric Markov Perfect equilibria in weakly undominated strategies are canonical.

C.2 Proof of Proposition 1-(ii)-(iii)-(iv)

Transition probabilities depend on one's perspective: either "objective" (i.e. the one of an outsider), or "subjective" (i.e. the one of a majority or minority member). This observation motivates our introducing the following notation: For any given group, we refer to the transition probability, say from group sizes i to j, *from a group member's perspective* as the probability that the group's size goes from i to j *conditional on this group member being still a member next period.*

For regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$, let $p^{\mathrm{r}}_{i,j}$ be the transition probability from a majority member's perspective, i.e. the probability that the majority size moves from $i \geq k$ to $j \in \{i-1, i, i+1\}$[78] (note that $p^{\mathrm{r}}_{i,j} = 0$ if $|i-j| > 1$) from one period to another conditional on the majority member still being in the organization in the following period[79] (which has probability $(N-1)/N$). Then, for any $M > k$ and in the entrenched equilibrium ($\mathrm{r} = \mathrm{e}$):

$$\begin{cases} p^{\mathrm{e}}_{M,M+1} = (1-x)\left(1 - \dfrac{M+1}{N}\right)\dfrac{N}{N-1} = (1-x)\left(1 - \dfrac{M}{N-1}\right) \\ p^{\mathrm{e}}_{M,M} = \left[(1-x)\dfrac{M}{N} + x\left(1 - \dfrac{M}{N}\right)\right]\dfrac{N}{N-1} = (1-x)\dfrac{M}{N-1} + x\left(1 - \dfrac{M-1}{N-1}\right) \\ p^{\mathrm{e}}_{M,M-1} = x\dfrac{M-1}{N}\dfrac{N}{N-1} = x\dfrac{M-1}{N-1} \end{cases} \tag{17}$$

[78] If $j = k-1$, then the majority becomes the minority and the new majority is of size k.

[79] Consistently with our notation throughout the paper, the conditioning on the majority member still being in the organization in the following period makes the relevant discount factor be δ, i.e. the life-adjusted discount factor. We could have equivalently written the *unconditioned* transition probabilities, i.e. the probability of majority size going from i to j *and* the majority member still being in the organization in the following period, which would have led to using the pure-time discount δ_0.

and

$$\begin{cases} p^{\mathrm{e}}_{k,k+1} = \left(1 - \dfrac{k+1}{N}\right)\dfrac{N}{N-1} = 1 - \dfrac{k}{N-1} \\ p^{\mathrm{e}}_{k,k} = \dfrac{k}{N}\dfrac{N}{N-1} = \dfrac{k}{N-1} \\ p^{\mathrm{e}}_{k,k-1} = 0 \end{cases} \tag{18}$$

For any $i, j \in \{1, ..., N-1\}$ and $t \in \mathbb{N}_+$, let $\pi^{\mathrm{e}}_{i,j}(t)$ be the t-period transition probability from i to j in the entrenched equilibrium from a majority member's perspective[80]. In other words, $\pi^{\mathrm{e}}_{i,j}(t)$ is the probability that starting from i, the majority size is equal to j after t periods conditional on the majority member still being in the organization[81]. Hence, for any $i \in \{k, ..., N-1\}$ and $t \geq 0$,

$$\pi^{\mathrm{e}}_{i,M}(t+1) = p^{\mathrm{e}}_{M-1,M}\pi^{\mathrm{e}}_{i,M-1}(t) + p^{\mathrm{e}}_{M,M}\pi^{\mathrm{e}}_{i,M}(t) + p^{\mathrm{e}}_{M+1,M}\pi^{\mathrm{e}}_{i,M+1}(t)$$

We similarly explicit the transition probabilities from the perspective of a minority member. Let $\hat{p}^{\mathrm{e}}_{i,j}$ be the transition probability from a minority member's perspective, i.e. the probability that the majority size moves from $i \geq k$ to j from one period to another conditional on the minority member still being in the organization in the following period (which has probability $(N-1)/N$). Then, for any $M > k$ and in the entrenched equilibrium:

$$\begin{cases} \hat{p}^{\mathrm{e}}_{M,M+1} = (1-x)\left(1 - \dfrac{M+2}{N}\right)\dfrac{N}{N-1} = (1-x)\left(1 - \dfrac{M+1}{N-1}\right) \\ \hat{p}^{\mathrm{e}}_{M,M} = \left[(1-x)\dfrac{M+1}{N} + x\left(1 - \dfrac{M+1}{N}\right)\right]\dfrac{N}{N-1} = (1-x)\dfrac{M+1}{N-1} + x\left(1 - \dfrac{M}{N-1}\right) \\ \hat{p}^{\mathrm{e}}_{M,M-1} = x\dfrac{M}{N}\dfrac{N}{N-1} = x\dfrac{M}{N-1} \end{cases} \tag{19}$$

and

$$\begin{cases} \hat{p}^{\mathrm{e}}_{k,k+1} = \left(1 - \dfrac{k+2}{N}\right)\dfrac{N}{N-1} = 1 - \dfrac{k+1}{N-1} \\ \hat{p}^{\mathrm{e}}_{k,k} = \dfrac{k+1}{N}\dfrac{N}{N-1} = \dfrac{k+1}{N-1} \\ \hat{p}^{\mathrm{e}}_{k,k-1} = 0 \end{cases} \tag{20}$$

For any $i, j \in \{1, ..., N-1\}$, and $t \in \mathbb{N}_+$, let $\hat{\pi}^{\mathrm{e}}_{i,j}(t)$ be the t-period transition probability from i to j in the entrenchment equilibrium from a minority member's perspective. In other words, $\hat{\pi}^{\mathrm{e}}_{i,j}(t)$ is the probability that starting from i, the minority size is equal to j after t periods conditional on the minority member still being in the organization. Hence, for any $i \in \{k, ..., N-1\}$ and $t \geq 0$,

$$\hat{\pi}^{\mathrm{e}}_{i,M}(t+1) = \hat{p}^{\mathrm{e}}_{M-1,M}\hat{\pi}^{\mathrm{e}}_{i,M-1}(t) + \hat{p}^{\mathrm{e}}_{M,M}\hat{\pi}^{\mathrm{e}}_{i,M}(t) + \hat{p}^{\mathrm{e}}_{M+1,M}\hat{\pi}^{\mathrm{e}}_{i,M+1}(t)$$

For the meritocratic equilibrium, transition probabilities are given by (17) for majority members, and by (19) for minority members.

[80] We focus on the entrenched equilibrium for the sake of exposition. Transition probabilities in the meritocratic regime have a more complex expression, with $\pi^{\mathrm{m}}_{i,j}(t)$ defined as the t-period transition probability from i to j *from the perspective of a member of the group initially of size* i.

[81] So in particular $\pi^{\mathrm{e}}_{i,i}(0) = 1$ and $\pi^{\mathrm{e}}_{i,j}(0) = 0$ for any $j \neq i$.

For the meritocratic equilibrium, transition probabilities are given accordingly, depending on one's perspective.[82]

Note that because probabilities sum to 1,

$$\begin{cases} \left(\sum_{i=k+1}^{N-1} \hat{\pi}^{\mathrm{e}}_{k,i}(t)\right) - \left(\sum_{i=k+1}^{N-1} \pi^{\mathrm{e}}_{k+1,i}(t)\right) = -\left(\hat{\pi}^{\mathrm{e}}_{k,k}(t) - \pi^{\mathrm{e}}_{k+1,k}(t)\right) \\ \left(\sum_{i=k}^{N-1} \pi^{\mathrm{m}}_{k+1,i}(t)\right) - \left(\sum_{i=k}^{N-1} \pi^{\mathrm{m}}_{k-1,i}(t)\right) = -\left[\left(\sum_{i=1}^{k-1} \pi^{\mathrm{m}}_{k+1,i}(t)\right) - \left(\sum_{i=1}^{k-1} \pi^{\mathrm{m}}_{k-1,i}(t)\right)\right] \end{cases} \tag{21}$$

Proof of claims (ii) and (iii). We now turn to the statement of the existence result. Because the majority's choice when the majority candidate brings higher value than (or the same value as) the minority one is a no-brainer, let us examine the case in which the majority is tight and the minority candidate is more talented. The optimality decision hinges on the choice of the size and identity of the future majority.

Condition for existence of the meritocratic equilibrium. Leaving aside control considerations, choosing the less-deserving majority candidate when the majority is tight involves a cost $s - b$. To evaluate the impact of a potential switch of control, which will occur as we just saw with conditional probability $(k-1)/(N-1)$, note that in a meritocratic equilibrium, the present discounted expected quality of future appointees does not depend on the allocation of control. The only impact of the change in control is linked to homophily benefits when the two candidates have equal quality standing (which has probability $1-2x$), as control allows one to select the group's candidate. The present discounted probability of exercising control in future periods is higher if the majority keeps control next period than if it surrenders it. So overall a necessary condition of existence of a meritocratic equilibrium is:

$$s - b \geq \delta \frac{k-1}{N-1}(1-2x)b \sum_{t=0}^{+\infty} \delta^t \left[\left(\sum_{i=k}^{N-1} \pi^{\mathrm{m}}_{k+1,i}(t)\right) - \left(\sum_{i=k}^{N-1} \pi^{\mathrm{m}}_{k-1,i}(t)\right)\right]$$

And so the meritocratic equilibrium exists only if

$$\frac{s}{b} \geq \rho^{\mathrm{m}} \equiv 1 + \delta \frac{k-1}{N-1}(1-2x) \sum_{t=0}^{+\infty} \delta^t \left[\left(\sum_{i=k}^{N-1} \pi^{\mathrm{m}}_{k+1,i}(t)\right) - \left(\sum_{i=k}^{N-1} \pi^{\mathrm{m}}_{k-1,i}(t)\right)\right]$$

Lemma 3 implies that this condition is in fact also sufficient: as is intuitive, deviations from meritocracy are less appealing further away from a tight majority size, i.e. from immediate control considerations.

Condition for existence of the entrenched equilibrium. Again, choosing the less talented majority candidate yields a direct payoff loss $s - b$. Then, with probability $(k-1)/(N-1)$, the surrendering of control translates into a permanent loss of homophily benefits whenever the two candidates have equal quality standing, which has probability $1-2x$. This cost is equal to

$$\frac{\delta}{1-\delta}(1-2x)b$$

[82] For instance, transition probabilities for $M > k$ are given by (17) for majority members, and by (19) for minority members.

Moreover, because the new majority will itself be entrenched, i.e. always voting for its own candidate whenever the majority is tight, the surrendering of control entails a further additional loss of homophily benefit proportional to $2xb$ whenever the majority is tight, along with the difference in homophily benefits associated with meritocratic decisions, i.e. choosing a talented minority candidate instead of an untalented majority candidate, at any majority size $M \geq k+1$. The latter would seem unwarranted as the two groups then agree on the decision to pick the more talented candidate; its existence comes from the fact that transition probabilities depend on one's perspective. Put together, these two terms add up to

$$\delta \frac{k-1}{N-1} 2xb \sum_{t=0}^{+\infty} \delta^t \pi^{\mathrm{e}}_{k+1,k}(t) - \delta \frac{k-1}{N-1} xb \sum_{t=0}^{+\infty} \delta^t \left[\left(\sum_{i=k+1}^{N-1} \hat{\pi}^{\mathrm{e}}_{k,i}(t) \right) - \left(\sum_{i=k+1}^{N-1} \pi^{\mathrm{e}}_{k+1,i}(t) \right) \right]$$

Another way to interpret the homophily payoff terms consists in noticing that the expected per-period payoff of a majority (resp. minority) member is equal to $(1-x)b$ (resp. xb) whenever the majority is not tight ($M \geq k+1$), while it is equal to b (resp. 0) when majority is tight ($M = k$).

Finally, again because the new majority is itself entrenched, and since the shift in control implies that perspectives change, the surrendering of control yields a quality payoff equal to

$$\begin{aligned} &\delta \frac{k-1}{N-1} (\overline{x} + x) s \sum_{t=0}^{+\infty} \delta^t \left[\left(\sum_{i=k+1}^{N-1} \hat{\pi}^{\mathrm{e}}_{k,i}(t) \right) - \left(\sum_{i=k+1}^{N-1} \pi^{\mathrm{e}}_{k+1,i}(t) \right) \right] \\ &\quad + \delta \frac{k-1}{N-1} \overline{x} s \sum_{t=0}^{+\infty} \delta^t \left(\hat{\pi}^{\mathrm{e}}_{k,k}(t) - \pi^{\mathrm{e}}_{k+1,k}(t) \right) \end{aligned}$$

So overall a necessary condition for the existence of an entrenched equilibrium is

$$\begin{aligned} b - s \geq\ & \delta \frac{k-1}{N-1} (\overline{x} + x) s \sum_{t=0}^{+\infty} \delta^t \left[\left(\sum_{i=k+1}^{N-1} \hat{\pi}^{\mathrm{e}}_{k,i}(t) \right) - \left(\sum_{i=k+1}^{N-1} \pi^{\mathrm{e}}_{k+1,i}(t) \right) \right] \\ &+ \delta \frac{k-1}{N-1} \overline{x} s \sum_{t=0}^{+\infty} \delta^t \left(\hat{\pi}^{\mathrm{e}}_{k,k}(t) - \pi^{\mathrm{e}}_{k+1,k}(t) \right) - \frac{k-1}{N-1} \frac{\delta}{1-\delta} (1-2x) b \\ &- \delta \frac{k-1}{N-1} 2xb \sum_{t=0}^{+\infty} \delta^t \pi^{\mathrm{e}}_{k+1,k}(t) + \delta \frac{k-1}{N-1} xb \sum_{t=0}^{+\infty} \delta^t \left[\left(\sum_{i=k+1}^{N-1} \hat{\pi}^{\mathrm{e}}_{k,i}(t) \right) - \left(\sum_{i=k+1}^{N-1} \pi^{\mathrm{e}}_{k+1,i}(t) \right) \right] \end{aligned}$$

Let (22) be the inequality:

$$1 + \delta \frac{k-1}{N-1} x \sum_{t=0}^{+\infty} \delta^t \left(\pi^{\mathrm{e}}_{k+1,k}(t) - \hat{\pi}^{\mathrm{e}}_{k,k}(t) \right) > 0. \tag{22}$$

Define ρ^{e} as

$$\rho^{\mathrm{e}} \equiv \begin{cases} \dfrac{1 + \dfrac{k-1}{N-1} \dfrac{\delta}{1-\delta} (1-2x) + \delta \dfrac{k-1}{N-1} x \sum_{t=0}^{+\infty} \delta^t \left(\pi^{\mathrm{e}}_{k+1,k}(t) + \hat{\pi}^{\mathrm{e}}_{k,k}(t) \right)}{1 + \delta \dfrac{k-1}{N-1} x \sum_{t=0}^{+\infty} \delta^t \left(\pi^{\mathrm{e}}_{k+1,k}(t) - \hat{\pi}^{\mathrm{e}}_{k,k}(t) \right)} & \text{if (22) holds,} \\ +\infty \qquad \text{otherwise.} \end{cases}$$

Then the above argument suggests that the entrenched equilibrium exists only if $s/b \leq \rho^{\mathrm{e}}$. As the series

term in (22) is negative for all t (see Lemma 4 below), there might exist an entrenched equilibrium for all values of s and b (and in particular for $b = 0$) for δ sufficiently close to 1, and thus we set $\rho^{\mathrm{e}} = +\infty$. Nonetheless, for a positive rate of time preference (which we assumed) – i.e. $\delta < (N-1)/N$ –, the entrenched equilibrium exists only on a finite interval: $\rho^{\mathrm{e}} < +\infty$.[83]

As hinted above, Lemma 3 yields that these necessary conditions are also sufficient for these equilibria to exist. Hence, the entrenched (resp. meritocratic) equilibrium exists if and only if $s/b \leq \rho^{\mathrm{e}}$ (resp. $s/b \geq \rho^{\mathrm{m}}$).

Lastly, we show that the bounds ρ^{e} and ρ^{m} satisfy the following inequalities:[84]

$$1 \leq 1 + \delta \frac{k-1}{N-1}(1-2x) \leq \rho^{\mathrm{m}} \leq 1 + \frac{\delta}{1-\delta}\frac{k-1}{N-1}(1-2x) < \rho^{\mathrm{e}} < +\infty \tag{23}$$

The upper and lower bounds on ρ^{m} may be decomposed as follows: $(1-2x)$ is the probability of a homophily benefit from control, $(k-1)/(N-1)$ the (conditional) probability of losing the majority when its end-of-period size is k, while δ (resp. $\delta/(1-\delta)$) are the time-discounted weights corresponding to a transient (resp. permanent) loss of control.[85]

The bounds on ρ^{e} and ρ^{m} in Inequality (23) derive from the following lemma.

Lemma 4. *For all $t \geq 0$,*

(i) $\pi^{e}_{k+1,k}(t) \leq \hat{\pi}^{e}_{k,k}(t)$

(ii) $\sum_{i \geq k} \pi^{m}_{k+1,i}(t) \geq \sum_{i \geq k} \pi^{m}_{k-1,i}(t)$

Proof. We use a result relying on the properties of monotone Markov chains (see Daley 1968, Keilson-Kester 1977).
(i) Let P (resp. $\hat{P}$) be the stochastic matrix associated with the process $M(t)$ (resp. $\hat{M}(t)$) defined as the probability distribution over majority sizes $\{k, ..., N-1\}$ from a majority (resp. minority) member's perspective[86]. Namely, for any $i, j \in \{1, ..., k\}$,

$$P_{ij} = p^{\mathrm{e}}_{k+i-1,k+j-1}, \qquad \text{and} \qquad \hat{P}_{ij} = \hat{p}^{\mathrm{e}}_{k+i-1,k+j-1}$$

We first note that both P and $\hat{P}$ are *(strictly) stochastically-monotone* as $P_{i\cdot}$ stochastically dominates $P_{i'\cdot}$ whenever $i > i'$ (and similarly for $\hat{P}$)[87]. We then note that P and $\hat{P}$ are stochastically comparable with $P_{i\cdot}$ stochastically dominating $\hat{P}_{i\cdot}$ for any $i \in \{1, ..., k\}$. Furthermore, the process $M(t)$ starts from the initial state $M(0) = (0, 1, 0, ...)$ which stochastically dominates the initial state of the process $\hat{M}(t)$,

[83] See Section C.3 for the proof of this result.

[84] The proof that $\rho^{\mathrm{e}} < +\infty$ is delayed to Section C.3.

[85] Note that ρ^{m} reaches its upper bound as x goes to 0. In the limit, it is equal to $1 + \frac{\delta}{1-\delta}\frac{k-1}{N-1}$, which is intuitive: the majority weights the current-period payoff $s-b$ against the constant homophily loss in future periods due to the permanent loss of control (times its probability of occurrence $(k-1)/(N-1)$).

[86] The i-th component of $M(t)$ is the probability (from the perspective of a majority member) that the majority be of size $k+1-i$ at period t. In particular, if at time 0 the majority is known to have size $k+1$, then $M(0) = (0, 1, 0, ..., 0)$, and at any later time t, $M(t) = \left(\pi^{\mathrm{e}}_{k+1,k}(t), ..., \pi^{\mathrm{e}}_{k+1,N-1}(t)\right)$. Similarly, if at time 0 the majority is known to have size k, then $\hat{M}(0) = (1, 0, , ..., 0)$, and at any later time t, $\hat{M}(t) = \left(\hat{\pi}^{\mathrm{e}}_{k,k}(t), ..., \hat{\pi}^{\mathrm{e}}_{k,N-1}(t)\right)$.

[87] Namely, for any $j^* \in \{1, ..., k\}$, $\Sigma_{j \geq j^*} P_{ij} \geq \Sigma_{j \geq j^*} P_{i'j}$.

that is $\hat{M}(0) = (1, 0, , 0, ...)$.

Hence, a standard argument implies that for any $t > 0$, the distribution $M(t)$ stochastically dominates the distribution $\hat{M}(t)$ [88]. In particular, we have that for any $t > 0$,

$$\sum_{i=k+1}^{N-1} \pi^{\mathrm{e}}_{k+1,i}(t) \geq \sum_{i=k+1}^{N-1} \hat{\pi}^{\mathrm{e}}_{k,i}(t),$$

which, since probabilities sum to 1, is equivalent to: $\pi^{\mathrm{e}}_{k+1,k}(t) \leq \hat{\pi}^{\mathrm{e}}_{k,k}(t)$.

(ii) In order to establish the lower bound on ρ^{m} and thus Inequality (23), we note that:

$$\left(\sum_{i\geq k} \pi^{\mathrm{m}}_{k+1,i}(t)\right) - \left(\sum_{i\geq k} \pi^{\mathrm{m}}_{k-1,i}(t)\right) > 0 \qquad \forall t \geq 0$$

This inequality can be shown with the same technique as the one used in the proof of claim (i) by considering the process of one's successive in-group sizes in the meritocratic equilibrium, either starting from the initial state $k+1$ or $k-1$. Indeed, the same conditions are satisfied, as (a) both processes (of probability distribution over one's successive in-group sizes) share the same transition matrix[89] which is stochastically monotone, and (b) the initial state with mass 1 in $k+1$ stochastically dominates the initial state with mass 1 in $k-1$. Hence the stochastic-comparison argument applies, yielding that the process of one's in-group size starting from $k+1$ stochastically dominates at any time $t \geq 0$ the process starting from $k-1$, and thus in particular,

$$\sum_{i\geq k} \pi^{\mathrm{m}}_{k+1,i}(t) > \sum_{i\geq k} \pi^{\mathrm{m}}_{k-1,i}(t)$$

□

Proof of claim (iii). The result derives from the explicit expressions of the existence thresholds together with Lemma 4. Indeed, by the proof of Lemma 4 and Proposition 1, we have that for all $t \geq 0$,

$$\pi^{\mathrm{e}}_{k+1,k}(t) - \hat{\pi}^{\mathrm{e}}_{k,k}(t) \leq 0, \qquad \text{and} \qquad \left(\sum_{i\geq k} \pi^{\mathrm{m}}_{k+1,i}(t)\right) - \left(\sum_{i\geq k} \pi^{\mathrm{m}}_{k-1,i}(t)\right) \geq 0$$

Using term-by-term differentiation of the series yields the result: $\partial\rho^{\mathrm{m}}/\partial\delta \geq 0, \partial\rho^{\mathrm{e}}/\partial\delta \geq 0$ for all $\delta \in [0, 1)$. Moreover, using term-by-term differentiation of the series for ρ^{m} and explicit computations for ρ^{e} yields

$$\left.\frac{\partial\rho^{\mathrm{m}}}{\partial\delta}\right|_{\delta=0} = \frac{k-1}{N-1}(1-2x) \qquad \text{and} \qquad \left.\frac{\partial\rho^{\mathrm{e}}}{\partial\delta}\right|_{\delta=0} = \frac{k-1}{N-1}$$

[88] A sketch of the proof is as follows. Proceed by induction on t. The result for $t = 0$ holds as noted in the text. Suppose that $M(t)$ stochastically dominates $\hat{M}(t)$. Then, since P stochastically dominates $\hat{P}$, we have that $\hat{M}(t)P$ stochastically dominates $\hat{M}(t)\hat{P}$. Since P is stochastically-monotone, $M(t)P$ stochastically dominates $\hat{M}(t)P$. Thus, by transitivity, $M(t)P$ stochastically dominates $\hat{M}(t)\hat{P}$. In other words, $M(t+1)$ stochastically dominates $\hat{M}(t+1)$, which concludes the proof.

[89] Namely, the matrix P^{m} with components $P_{ij} = p^{\mathrm{m}}_{i,j}$ for any $i, j \in \{1, ..., N-1\}$.

C.3 Proof of Proposition 1: Entrenchment exists only on a finite interval ($\rho^{\mathrm{e}} < \infty$)

We show in this section that $\rho^{\mathrm{e}} < \infty$ [90]. The result is trivial for $k = 2$ (e.g. using the explicit expression of ρ^{e}). Let $k \geq 3$. The result is obvious for $x = 0$ (again by the explicit expression of ρ^{e}). We thus consider the case $x > 0$.

Let V_i^{e} denote the value function in the entrenched equilibrium, and define as before $u_i^{\mathrm{e}} \equiv V_{i+1}^{\mathrm{e}} - V_i^{\mathrm{e}}$. Fix $s > 0$. For any $i \in \{1, ..., N-2\}$, u_i^{e} is clearly continuous with respect to $b \in [0, +\infty)$.

The (one-shot) deviation differential payoff from entrenchment to meritocracy in $M = k$ writes:

$$s - b + \delta \frac{k-1}{N-1}(V_{k-1} - V_{k+1}) = s - b - \delta \frac{k-1}{N-1}(u_{k-1}^{\mathrm{e}} + u_k^{\mathrm{e}})$$

Fix $b = 0$. If the above payoff is strictly positive for $b = 0$, then by continuity, it must be so on a neighbourhood of 0. Hence there exists $\overline{\rho} > 0$ such that for any $s/b > \overline{\rho}$, there exists a strictly profitable deviation from entrenchment to meritocracy, which yields the result: $\rho^{\mathrm{e}} < \infty$. We thus show that for $b = 0$:

$$s - \delta \frac{k-1}{N-1}(u_{k-1}^{\mathrm{e}} + u_k^{\mathrm{e}}) > 0 \tag{24}$$

We first show that the above inequality can be written as

$$\begin{aligned} &\frac{\delta x \frac{k-1}{N-1}}{1 - \delta x - \delta(1-x)\left[\frac{k+1}{N-1} + \frac{k-2}{N-1} a_{k+1}\right]} \\ &\quad \times \left(1 - \frac{\delta x \frac{k}{N-1}}{1 - \delta(1-x)\left[\frac{k}{N-1} + \frac{k-2}{N-1} b_{k+1}\right]} - \frac{\delta x \frac{k-2}{N-1}}{1 - \delta(1-x)\left[\frac{k+1}{N-1} + \frac{k-3}{N-1} b_{k+2}\right]}\right) < 1 \end{aligned} \tag{25}$$

where the vectors $(a_{k+l})_{l=1}^{k-2}$, $(b_{k+l})_{l=1}^{k-2}$ are defined recursively by

$$\begin{cases} a_{k+l} = \dfrac{\delta x \frac{k+l}{N-1}}{1 - \delta(1-x)\left[\frac{k+l+1}{N-1} + \frac{k-l-2}{N-1} a_{k+l+1}\right] - \delta x \frac{k-l-1}{N-1}} \\ a_{N-2} = \dfrac{\delta x \frac{N-2}{N-1}}{1 - \delta(1-x) - \delta \frac{x}{N-1}} \end{cases}$$

[90] The proof also yields that $\rho^{\mathrm{e}\dagger}|_{s^\dagger > b} < \infty$ (thus in particular for $x^\dagger \geq 1/2$).

and

$$\begin{cases} b_{k+l} = \dfrac{\delta x \dfrac{k+l-1}{N-1}}{1-\delta(1-x)\left[\dfrac{k+l}{N-1}+\dfrac{k-l-2}{N-1}b_{k+l+1}\right]-\delta x\dfrac{k-l-1}{N-1}} \\ b_{N-2} = \dfrac{\delta x \dfrac{N-3}{N-1}}{1-\delta(1-x)\dfrac{N-2}{N-1}-\delta\dfrac{x}{N-1}} \end{cases}$$

Indeed, computations using (8)-(9) and (10)-(12) for the entrenchment equilibrium, give that:

$$\begin{cases} \left[1-\delta(1-x)\dfrac{k+1}{N-1}-\delta x\right]\left(V^{\mathrm{e}}_{k+1}-V^{\mathrm{e}}_{k-1}\right) \\ \qquad = xs+\delta(1-x)\dfrac{k-2}{N-1}\left(V^{\mathrm{e}}_{k+2}-V^{\mathrm{e}}_{k-2}\right)-\delta x\dfrac{k}{N-1}u^{\mathrm{e}}_{k}+\delta x\dfrac{k-2}{N-1}u^{\mathrm{e}}_{k-2} \\ V^{\mathrm{e}}_{k+2}-V^{\mathrm{e}}_{k-2} = a_{k+1}\left(V^{\mathrm{e}}_{k+1}-V^{\mathrm{e}}_{k-1}\right) \\ u^{\mathrm{e}}_{k+1} = b_{k+1}u^{\mathrm{e}}_{k} \\ u^{\mathrm{e}}_{k-3} = b_{k+2}u^{\mathrm{e}}_{k-2} \end{cases}$$

and thus, by rearranging[91], (24) is equivalent to (25).

Remark. By construction, (a_{k+l}) and (b_{k+l}) are increasing with l, and for any l, $b_{k+l} < a_{k+l} < 1$. Moreover, for any l, a_{k+l} and b_{k+l} are increasing with respect to x and δ.[92]

We show that for any $x \in [0,1]$ and $\delta \in [0,(N-1)/N)$, inequality (25) is satisfied[93]. The above remark on the properties of $(a_{k+l})_{l=1}^{k-2}$, $(b_{k+l})_{l=1}^{k-2}$, yields that the term on the first line (resp. second line) is strictly increasing (resp. decreasing) with respect to x and δ. Moreover, by continuity, (25) is clearly satisfied for (x,δ) in a neighbourhood of $(0,\delta)$, $(x,0)$, and most interestingly of $(1,(N-1)/N)$.

[91] Using in particular that (8)-(9) imply:

$$\begin{cases} u^{\mathrm{e}}_{k} = xs+\delta(1-x)\left[\left(1-\dfrac{k+1}{N-1}\right)u^{\mathrm{e}}_{k+1}+\dfrac{k}{N-1}u^{\mathrm{e}}_{k}\right] \\ u^{\mathrm{e}}_{k-2} = -xs+\delta(1-x)\left[\dfrac{k+1}{N-1}u^{\mathrm{e}}_{k-2}+\left(1-\dfrac{k+2}{N-1}\right)u^{\mathrm{e}}_{k-3}\right] \end{cases}$$

[92] These results can be shown by downward induction starting from $l = N-2$.

[93] The case $x \geq 1/2$ is equivalent to the homogamic-evaluation-capability setting with $x^{\dagger} \geq 1/2$. Indeed, the homogamic-evaluation-capability equivalent of (24) is:

$$\frac{\delta x\dfrac{k-1}{N-1}}{1-\delta x^{\dagger}-\delta(1-x^{\dagger})\left[\dfrac{k+1}{N-1}+\dfrac{k-2}{N-1}a^{\dagger}_{k+1}\right]} \times\left(1-\frac{\delta x^{\dagger}\dfrac{k}{N-1}}{1-\delta(1-x^{\dagger})\left[\dfrac{k}{N-1}+\dfrac{k-2}{N-1}b^{\dagger}_{k+1}\right]}-\frac{\delta x^{\dagger}\dfrac{k-2}{N-1}}{1-\delta(1-x^{\dagger})\left[\dfrac{k+1}{N-1}+\dfrac{k-3}{N-1}b^{\dagger}_{k+2}\right]}\right)<\frac{x}{x^{\dagger}}$$

with the corresponding families $(a^{\dagger}_{k+l})_{l=1}^{k-2}$, $(b^{\dagger}_{k+l})_{l=1}^{k-2}$ defined as before by replacing x with $x^{\dagger}$.

Using the inequality $b_{k+1} < b_{k+2} < 1$, a sufficient condition for (25) to be satisfied is

$$\delta x \frac{k-1}{N-1}\left(1-\delta x\frac{N-2}{N-1}-\delta(1-x)\left[\frac{k}{N-1}+\frac{k-2}{N-1}b_{k+1}\right]\right) \tag{26}$$
$$/\left[\left(1-\delta(1-x)\left[\frac{k}{N-1}+\frac{k-2}{N-1}b_{k+1}\right]\right)\left(1-\delta x-\delta(1-x)\left[\frac{k+1}{N-1}+\frac{k-2}{N-1}a_{k+1}\right]\right)\right]<1$$

or equivalently,

$$\delta x \frac{k-1}{N-1}\left(1-\delta x\frac{N-2}{N-1}-\delta(1-x)\left[\frac{k}{N-1}+\frac{k-2}{N-1}b_{k+1}\right]\right) \tag{27}$$
$$-\left(1-\delta(1-x)\left[\frac{k}{N-1}+\frac{k-2}{N-1}b_{k+1}\right]\right)\left(1-\delta x-\delta(1-x)\left[\frac{k+1}{N-1}+\frac{k-2}{N-1}a_{k+1}\right]\right)<0$$

The above inequalities are strictly stronger than (25) for any $x \in (0,1)$, and coincide with (25) in $x = 1$. Moreover, it holds for any $(x, \delta) = (1, \delta)$ where $\delta \in [0, (N-1)/N]$.

We show that for any $x \in [0,1]$, (i) the LHS in (27) increases with δ over $[0,(N-1)/N]$, and (ii) this maximum (LHS with $\delta = (N-1)/N$) is strictly negative.

(i) In order to alleviate the notation, let C_a and C_b be defined as

$$C_a \equiv \frac{k+1}{N-1}+\frac{k-2}{N-1}a_{k+1}, \qquad \text{and} \qquad C_b \equiv \frac{k}{N-1}+\frac{k-2}{N-1}b_{k+1}$$

Since $b_{k+1} < a_{k+1} < 1$, then $C_b < C_a < 1$. The derivative of the LHS in (27) with respect to δ writes after rearranging:

$$\varphi(\delta) \equiv x\left(1+\frac{k-1}{N-1}\right)+(1-x)(C_a+C_b)-2\delta\left[x(1-x)\left(1+\frac{k-1}{N-1}\right)C_b+(1-x)^2C_aC_b+x^2\frac{k-1}{N-1}\frac{N-2}{N-1}\right]$$
$$+\delta(1-x)\frac{k-2}{N-1}\left(\frac{\partial a_{k+1}}{\partial\delta}\left[1-\delta(1-x)C_b\right]+\frac{\partial b_{k+1}}{\partial\delta}\left[1-\delta(1-x)C_a-\delta x\left(1+\frac{k-1}{N-1}\right)\right]\right)$$

Using a downward induction argument on the sequences $(a_{k+l})_l, (b_{k+l})_l$ yields that $\partial a_{k+1}/\partial\delta > \partial b_{k+1}/\partial\delta$.[94]

[94]The result follows from the observation that

$$\frac{\partial a_{N-2}}{\partial\delta}=\frac{x\frac{N-2}{N-1}}{\left(1-\delta(1-x)-\delta\frac{x}{N-1}\right)^2}>\frac{x\frac{N-3}{N-1}}{\left(1-\delta(1-x)\frac{N-2}{N-1}-\delta\frac{x}{N-1}\right)^2}=\frac{\partial b_{N-2}}{\partial\delta}$$

and for any $l \in \{1,...,k-3\}$,

$$\frac{\partial a_{k+l}}{\partial\delta}=\frac{x\frac{k+l}{N-1}+\delta^2x(1-x)\frac{k+l}{N-1}\frac{k-l-2}{N-1}\frac{\partial a_{k+l+1}}{\partial\delta}}{\left(1-\delta(1-x)\left[\frac{k+l+1}{N-1}+\frac{k-l-2}{N-1}a_{k+l+1}\right]-\delta x\frac{k-l-1}{N-1}\right)^2}$$
$$>\frac{x\frac{k+l-1}{N-1}+\delta^2x(1-x)\frac{k+l-1}{N-1}\frac{k-l-2}{N-1}\frac{\partial b_{k+l+1}}{\partial\delta}}{\left(1-\delta(1-x)\left[\frac{k+l}{N-1}+\frac{k-l-2}{N-1}b_{k+l+1}\right]-\delta x\frac{k-l-1}{N-1}\right)^2}=\frac{\partial b_{k+l}}{\partial\delta}$$

Therefore,

$$\begin{aligned}\phi(\delta) &\equiv \frac{\partial a_{k+1}}{\partial \delta}\left[1-\delta(1-x)C_b\right]+\frac{\partial b_{k+1}}{\partial \delta}\left[1-\delta(1-x)C_a-\delta x\left(1+\frac{k-1}{N-1}\right)\right]\\ &\geq \frac{\partial b_{k+1}}{\partial \delta}\left[2-\delta(1-x)(C_a+C_b)-\delta x\left(1+\frac{k-1}{N-1}\right)\right]>0\end{aligned}$$

Let $\psi(\delta)\equiv\varphi(\delta)-\delta(1-x)\frac{k-2}{N-1}\phi(\delta)$. Then, by rearranging,

$$\begin{aligned}\psi(\delta) &= x\left(1+\frac{k-1}{N-1}\right)+(1-x)(C_a+C_b)-2\delta\left[x(1-x)\left(1+\frac{k-1}{N-1}\right)C_b+(1-x)^2C_aC_b+x^2\frac{k-1}{N-1}\frac{N-2}{N-1}\right]\\ &= x\left[1+\frac{k-1}{N-1}-\delta(1-x)\left(1+\frac{k-1}{N-1}\right)C_b-\delta x\left(\frac{N-2}{N-1}\right)^2\right]\\ &\quad+(1-x)\left[\left(C_a-\delta xC_b-\delta(1-x)C_aC_b\right)+\left(C_b-\delta x\frac{k-1}{N-1}C_b-\delta(1-x)C_aC_b\right)\right]\geq 0\end{aligned}$$

where the last inequality stems from the fact that $k/(N-1)<C_b<C_a<1$. Hence $\varphi(\delta)>0$ for any $x\in[0,1]$. Consequently, the LHS in (27) is strictly increasing with respect to δ, and thus reaches its maximum over $[0,(N-1)/N]$ in $\delta=(N-1)/N$.

(ii) We now let $\delta=(N-1)/N$ and show that the LHS in (27) with $\delta=(N-1)/N$ is strictly negative. The latter then writes as

$$\begin{aligned}LHS &\equiv x\frac{k-1}{N}\left(1-x\frac{N-2}{N}-(1-x)\left[\frac{k}{N}+\frac{k-2}{N}b_{k+1}\right]\right)\\ &\quad-\left(1-(1-x)\left[\frac{k}{N}+\frac{k-2}{N}b_{k+1}\right]\right)\left(1-x\frac{N-1}{N}-(1-x)\left[\frac{k+1}{N}+\frac{k-2}{N}a_{k+1}\right]\right)\\ &= x\frac{k-1}{N}\left(\frac{1}{N}+(1-x)\frac{k-2}{N}(a_{k+1}-b_{k+1})\right)\\ &\quad-\left(\frac{k+1}{N}-\frac{1-x}{N}-(1-x)\frac{k-2}{N}b_{k+1}\right)\left(1-x\frac{N-1}{N}-(1-x)\left[\frac{k+1}{N}+\frac{k-2}{N}a_{k+1}\right]\right)\end{aligned}$$

where b_{k+1} and a_{k+1} are evaluated in $\delta=(N-1)/N$. By using that $b_{k+1}<1$ and rearranging, we get that

$$\begin{aligned}LHS &\leq x\frac{k-1}{N}\left(\frac{1}{N}+(1-x)\frac{k-2}{N}(a_{k+1}-b_{k+1})\right)\\ &\quad-\left(\frac{2}{N}+x\frac{k-1}{N}\right)\left(1-x\frac{N-1}{N}-(1-x)\left[\frac{k+1}{N}+\frac{k-2}{N}a_{k+1}\right]\right)\\ &= -\frac{2}{N^2}-(1-x)\frac{2}{N}\frac{k-2}{N}\left[1-a_{k+1}\right]-x(1-x)\frac{k-1}{N}\frac{k-2}{N}\left[1-2a_{k+1}+b_{k+1}\right]\end{aligned}$$

Hence, a sufficient condition for the LHS in (27) to be strictly negative is that $1-2a_{k+1}+b_{k+1}>0$. This actually holds, which concludes the proof: it can in fact be shown that for any $l\in\{1,...,k-2\}$, $1-2a_{k+l}+b_{k+l}\geq 0$ (with strict inequality whenever $x<1$).[95]

[95] The argument is as follows. One first notes that since for any $l\in\{1,...,k-2\}$, $\partial a_{k+l}/\partial\delta\geq\partial b_{k+l}/\partial\delta>0$, the term $[1-2a_{k+l}+b_{k+l}]$ is strictly bounded below by its value for $\delta=(N-1)/N$. The rest of the argument derives from downward

D Proof of Proposition 2

The result for minority members is obvious. We show the one for majority members. For any $i \in \{k, ..., N-1\}$, let $v_i \equiv V_i^{\mathrm{m}} - V_i^{\mathrm{e}}$. Note that by construction, for any $i \geq k+1$,

$$v_i = \delta(1-x)\left[\frac{i}{N-1}v_i + \left(1 - \frac{i}{N-1}\right)v_{i+1}\right] + \delta x\left[\frac{i-1}{N-1}v_{i-1} + \left(1 - \frac{i-1}{N-1}\right)v_i\right]$$

and therefore,

$$\left[1 - \delta(1-x)\frac{i}{N-1} - \delta x\left(1 - \frac{i-1}{N-1}\right)\right]v_i = \delta(1-x)\left(1 - \frac{i}{N-1}\right)v_{i+1} + \delta x\frac{i-1}{N-1}v_{i-1}, \tag{28}$$

while for $i = k$,

$$v_k = \Delta + \delta\left[\frac{k}{N-1}v_k + \left(1 - \frac{k}{N-1}\right)v_{k+1}\right]$$

where $\Delta \equiv x(s-b) + \delta x\frac{k-1}{N-1}\left(V_{k-1}^{\mathrm{m}} - V_{k+1}^{\mathrm{m}}\right) \geq 0$ – this last inequality stems from the definition of the meritocratic equilibrium and from the proof of Lemma 1: it is strict whenever $x > 0$ and either $b > 0$ or $s > b$ –, and thus

$$\left[1 - \delta\frac{k}{N-1}\right]v_k = \Delta + \delta\left(1 - \frac{k}{N-1}\right)v_{k+1} \tag{29}$$

Assume by contradiction that $v_{N-1} < 0$. Then, Equation (28) for $i = N-1$ implies that $v_{N-2} < v_{N-1} < 0$, and thus by induction that $v_k < v_{k+1} < ... < v_{N-1} < 0$. However, Equation (29) then yields

$$0 > (1-\delta)v_k > \Delta \geq 0,$$

induction showing the result for any l with $\delta = (N-1)/N$. Explicit computations yield that for $\delta = (N-1)/N$,

$$\left[1 - 2a_{N-2} + b_{N-2}\right] = \frac{(1-x)\frac{2}{N^2}}{\left(1 - (1-x)\frac{N-1}{N} - \frac{x}{N}\right)\left(1 - (1-x)\frac{N-2}{N} - \frac{x}{N}\right)} \geq 0$$

Then, for any $l \in \{1, ..., k-3\}$, the term $[1 - 2a_{k+l} + b_{k+l}]$ with $\delta = (N-1)/N$ has the same sign as

$$\begin{aligned}
&\left(1 - (1-x)\left[\frac{k+l+1}{N} + \frac{k-l-2}{N}a_{k+l+1}\right] - x\frac{k-l-1}{N}\right)\left(1 - (1-x)\left[\frac{k+l}{N} + \frac{k-l-2}{N}b_{k+l+1}\right] - x\frac{k-l-1}{N}\right) \\
&\quad - 2x\frac{k+l}{N}\left(1 - (1-x)\left[\frac{k+l}{N} + \frac{k-l-2}{N}b_{k+l+1}\right] - x\frac{k-l-1}{N}\right) \\
&\quad + x\frac{k+l-1}{N}\left(1 - (1-x)\left[\frac{k+l+1}{N} + \frac{k-l-2}{N}a_{k+l+1}\right] - x\frac{k-l-1}{N}\right) \\
&= (1-x)\left[\frac{k-l-1}{N} - \frac{k-l-2}{N}a_{k+l+1}\right]\left[\frac{k-l}{N} - x\frac{k-l-2}{N} - (1-x)\frac{k-l-2}{N}b_{k+l+1}\right] + x(1-x)\frac{k+l-1}{N}\frac{2}{N} \\
&\quad + x(1-x)\frac{k+l-1}{N}\frac{k-l-2}{N}\left[1 - 2a_{k+l+1} + b_{k+l+1}\right] \\
&\geq x(1-x)\frac{k+l-1}{N}\frac{k-l-2}{N}\left[1 - 2a_{k+l+1} + b_{k+l+1}\right]
\end{aligned}$$

which is a contradiction. Hence, $v_{N-1} \geq 0$, and by induction using Equation (28), $v_k \geq v_{k+1} \geq ... \geq v_{N-1} \geq 0$. In other words, for any $i \in \{k, ..., N-1\}$,

$$V_i^{\mathrm{m}} - V_i^{\mathrm{e}} \geq 0,$$

which concludes the proof. Note that the inequalities are strict whenever $\Delta > 0$, i.e. whenever $b > 0$ and $x > 0$ (with $s > b$ if $x = 1/2$). Moreover, the gap between the value functions in the two equilibria, $V_i^{\mathrm{m}} - V_i^{\mathrm{e}}$, decreases as the majority size moves further away from $M = k$.

E Proof of Lemma 2

We show successively that:

(i) $\nu_k^{\mathrm{e}} = 0$

(ii) for any $i \geq k+1$, we have that: $\dfrac{\nu_{i+1}^{\mathrm{e}}}{\nu_i^{\mathrm{e}}} = \dfrac{\nu_{i+1}^{\mathrm{m}}}{\nu_i^{\mathrm{m}}} = \dfrac{1-x}{x}\dfrac{N-i}{i+1}$,

(iii) $\nu_k^{\mathrm{e}} + \nu_{k+1}^{\mathrm{e}} < \nu_k^{\mathrm{m}} + \nu_{k+1}^{\mathrm{m}}$

and so, that the probability distribution $\{\nu_i^{\mathrm{e}}\}$ strictly first-order stochastically dominates $\{\nu_i^{\mathrm{m}}\}$.

Claim *(i)* derives from the fact that i refers to the size of the majority at the end of the period $i \in \{k, ..., 2k\}$. Note that in regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$,

$$\nu_N^{\mathrm{r}} = (1-x)\nu_N^{\mathrm{r}} + \frac{1-x}{N}\nu_{N-1}^{\mathrm{r}}$$

$$\text{and for } k+2 \leq i < N, \qquad \nu_i^{\mathrm{r}} = (1-x)\frac{N-(i-1)}{N}\nu_{i-1}^{\mathrm{r}} + \left[(1-x)\frac{i}{N} + x\frac{N-i}{N}\right]\nu_i^{\mathrm{r}} + x\frac{i+1}{N}\nu_{i+1}^{\mathrm{r}}$$

Claim *(ii)* follows by backward induction starting from $i = N$ and going down until $k+2$ included. Note that the explicit expression of the ergodic distribution in the entrenched equilibrium obtains with claims *(i)* and *(ii)* by writing $\Sigma_{i=k+1}^{N}\nu_i^{\mathrm{e}} = 1$. The explicit expression of the ergodic distribution in the meritocratic equilibrium obtains similarly noting that $(1-x)N\nu_k^{\mathrm{m}} = x(k+1)\nu_{k+1}^{\mathrm{m}}$. One has in particular that

$$\begin{cases} \nu_{k+1}^{\mathrm{m}}\left[\dfrac{x}{1-x}\dfrac{k+1}{N} + 1 + \displaystyle\sum_{i=1}^{k-1}\left(\frac{1-x}{x}\right)^i \prod_{l=1}^{i}\frac{k-l}{k+1+l}\right] = 1 \\ \nu_{k+1}^{\mathrm{e}}\left[1 + \displaystyle\sum_{i=1}^{k-1}\left(\frac{1-x}{x}\right)^i \prod_{l=1}^{i}\frac{k-l}{k+1+l}\right] = 1 \end{cases}$$

Lastly, claims *(i)* and *(ii)* together imply claim *(iii)*.

F Proof of Proposition 3

Let ρ^W be uniquely defined by

$$qN(N-1)\Big[1+\frac{x}{1-x}\frac{k+1}{N}+\sum_{l=1}^{k-1}\Big(\frac{1-x}{x}\Big)^l\prod_{j=1}^{l}\frac{k-j}{k+1+j}\Big]\rho^W$$
$$=\frac{2}{1-x}\Big[1+\sum_{l=1}^{k-1}(l+1)^2\Big(\frac{1-x}{x}\Big)^l\prod_{j=1}^{l}\frac{k-j}{k+1+j}\Big]$$

We show that $W^{\mathrm{m}}\geq W^{\mathrm{e}}$ if and only if $s/b\geq\rho^W$. The result thus obtains by showing that $\rho^W<1$ for all parameter values.

We first establish the explicit expression of ρ^W. By construction, we have that

$$B^{\mathrm{m}}-B^{\mathrm{e}}=\sum_{i=k}^{N}\big(\nu_i^{\mathrm{m}}-\nu_i^{\mathrm{e}}\big)\Big[i(i-1)+(N-i)(N-i-1)\Big]\tilde{b}$$

Hence, computations using the explicit expressions of the ergodic distributions (see Section E above) yield after rearranging:

$$\Big[\frac{x}{1-x}\frac{k+1}{N}+1+\sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\Big]\Big[1+\sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\Big]\big(B^{\mathrm{m}}-B^{\mathrm{e}}\big)$$
$$=-\frac{2x}{1-x}\frac{k+1}{N}\Big[1+\sum_{l=1}^{k-1}(l+1)^2\Big(\frac{1-x}{x}\Big)^l\prod_{j=1}^{l}\frac{k-j}{k+1+j}\Big]\tilde{b}$$

Similar computations for $(S^{\mathrm{m}}-S^{\mathrm{e}})$ yield:

$$\Big[\frac{x}{1-x}\frac{k+1}{N}+1+\sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\Big]\Big[1+\sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\Big]\big(S^{\mathrm{m}}-S^{\mathrm{e}}\big)$$
$$=N(N-1)x\frac{k+1}{N}\Big[\frac{x}{1-x}\frac{k+1}{N}+1+\sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\Big]\tilde{s}$$

The expression of ρ^W follows. Lastly, the inequality $\rho^W<1$ derives from the observations that for any $x\in[0,1/2]$, $N(N-1)>2(l+1)^2/(1-x)$ for any $l\leq k-2$, and that[96]

$$N(N-1)\Big[1+\Big(\frac{1-x}{x}\Big)^{k-1}\prod_{l=1}^{k-1}\frac{k-l}{k+1+l}\Big]>\frac{2}{1-x}\Big[1+k^2\Big(\frac{1-x}{x}\Big)^{k-1}\prod_{l=1}^{k-1}\frac{k-l}{k+1+l}\Big]$$

[96] Indeed, as the inequality $N(N-1)<2k^2/(1-x)$ holds if and only if $x>(k-1)/(N-1)$, we have that for any $x\in[0,1/2]$, the difference between the LHS minus the RHS is bounded below by

$$N(N-1)\Big[1+\Big(\frac{k}{k-1}\Big)^{k-1}\prod_{l=1}^{k-1}\frac{k-l}{k+1+l}\Big]-4\Big[1+k^2\Big(\frac{k}{k-1}\Big)^{k-1}\prod_{l=1}^{k-1}\frac{k-l}{k+1+l}\Big]>N(N-1)-4-N>0$$

where the first inequality derives from $\Big(\frac{k}{k-1}\Big)^{k-1}\prod_{l=1}^{k-1}\frac{k-l}{k+1+l}<1$, while the second holds for any $N\geq 4$.

G Proof of Proposition 4

We use a fixed-point argument to prove the existence of a class of equilibria characterized by a weakly decreasing decision rule $(\Delta_M)_M$ [97]. Let $\overline{u}$ be given by

$$\overline{u} \equiv \frac{1}{1-\delta}\Big(\mathbb{E}\big[(s+b)\mathbf{1}\{\hat{s}-s\leq b\}\big]+\mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s>b\}\big]\Big)$$

Note that $\big(\mathbb{E}\big[(s+b)\mathbf{1}\{\hat{s}-s\leq b\}\big]+\mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s>b\}\big]\big)$ is the highest flow payoff a majority member can guarantee, and consequently, $\overline{u}$ represents an upper bound on the majority's expected utility from a recruitment (i.e. its expected utility in the absence of control consideration). We define K as the set of sequences $(u_M)_{M\in\{k-1,...,N-2\}}$ such that (i) for any M, $u_M\in\big[0,\overline{u}\big]$ and (ii) the sequence $(u_M)_M$ is weakly decreasing. By construction, the set K is non-empty, compact and convex.

As earlier, let $\{V_i\}$ denote the value functions and $V\equiv(V_1,...,V_{N-1})$. For $i\in\{k-1,...,N-2\}$, let $u_i\equiv V_{i+1}-V_i$. In the equilibria we look for, whenever the majority has size $M\in\{k,...,N-1\}$, it favours a majority candidate with (discounted) talent s against a minority candidate with (discounted) talent $\hat{s}$ if and only if[98]

$$\hat{s}+\delta\Big[\frac{M-1}{N-1}V_{M-1}+\Big(1-\frac{M-1}{N-1}\Big)V_M\Big]\leq s+b+\delta\Big[\frac{M}{N-1}V_M+\Big(1-\frac{M}{N-1}\Big)V_{M+1}\Big],$$

i.e. if and only if

$$\hat{s}-s\leq b+\delta\Big[\frac{M-1}{N-1}u_{M-1}+\Big(1-\frac{M}{N-1}\Big)u_M\Big]$$

We denote by $\overline{s}\in[b,+\infty)$ the lowest real number such that $\mathbb{P}(\hat{s}-s\leq\overline{s})=1$ if it exists, and let $\overline{s}=+\infty$ otherwise. We first consider the "decision-rule" (cutoff) mapping $D:K\longrightarrow\big[0,\min(b+\delta\overline{u},\overline{s})\big]^k, u\longmapsto(D_M)_{M\in\{k,...,N-1\}}$, where

$$D_M(u)\equiv\begin{cases} b+\delta\Big[\frac{M-1}{N-1}u_{M-1}+\Big(1-\frac{M}{N-1}\Big)u_M\Big] & \text{if } b+\delta\Big[\frac{M-1}{N-1}u_{M-1}+\Big(1-\frac{M}{N-1}\Big)u_M\Big]<\overline{s}\\ \overline{s} & \text{otherwise}\end{cases}$$

Taking $V_{k-1}\geq 0$ as fixed, we consider the "value-function" mapping T defined as $T:\big[0,+\infty\big]^k\times\big[b,\overline{s}\big]^k\longrightarrow\big[0,+\infty\big]^k,\big((V_M)_M,(\Delta_M)_M\big)\longmapsto(T_M)_M$, where

$$\begin{aligned}T_M(V,\Delta)\equiv\mathbb{E}\big[(s+b)\mathbf{1}\{\hat{s}-s\leq\Delta_M\}\big]&+\delta\mathbb{P}(\hat{s}-s\leq\Delta_M)\Big[\frac{M}{N-1}V_M+\Big(1-\frac{M}{N-1}\Big)V_{M+1}\Big]\\&+\mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s>\Delta_M\}\big]+\delta\mathbb{P}(\hat{s}-s>\Delta_M)\Big[\frac{M-1}{N-1}V_{M-1}+\Big(1-\frac{M-1}{N-1}\Big)V_M\Big]\end{aligned}$$

[97] We thus focus on equilibria such that the decision rule only depends on the majority size.

[98] The assumption that ties are broken in favour of the majority candidate comes withouth loss of generality when vertical types are continuously distributed within each group.

In order to alleviate the notation, we define the functions h and h_1 as

$$\begin{cases} h(X) \equiv \mathbb{E}\big[(s+X)\mathbf{1}\{\hat{s}-s \leq X\}\big] + \mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s > X\}\big] \\ h_1(X) \equiv X - h(X) \end{cases}$$

Fix $V_{k-1} \geq 0$. For a sequence $u \equiv (u_M)_{M\in\{k-1,\dots,N-2\}} \in K$. We define the sequence $V(u) \equiv (V_M)_{M\in\{k,\dots,N-1\}}$ by upward induction by letting $V_M \equiv u_{M-1} + V_{M-1}$. Lastly, we define the mapping $\Upsilon : u \longmapsto \Upsilon(u)$ from K into itself by

$$\Upsilon_M(u) \equiv \min\Big\{T_{M+1}\big(V(u),D(u)\big) - T_M\big(V(u),D(u)\big), h(b)/(1-\delta)\Big\}$$

for any $M \in \{k-1,\dots,N-2\}$ (with the convention that $T_{k-1}\big(V(u),D(u)\big) \equiv V_{k-1}$). While bounding above $\Upsilon(u)$ is necessary to the argument, it does not threaten the existence of an equilibrium: indeed, $h(b)$ is the highest flow payoff (quality and homophily) that a majority member can guarantee[99]. Hence we have by construction that for any $u \in K$ and any $i \in \{k-1,\dots,N-2\}$, $\Upsilon_i(u) \leq \overline{u}$. With an abuse of notation, we omit in the following the min operator.

We now check that the mapping Υ is well-defined, i.e. that $\Upsilon(u) \in K$ for any $u \in K$.

$$\begin{aligned} T_M\big(V(u),D(u)\big) = \mathbb{E}\big[s\mathbf{1}\{\hat{s}-s \leq D_M(u)\}\big] + \mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s > D_M(u)\}\big] \\ + \mathbb{P}(\hat{s}-s \leq D_M(u))\left[b + \delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right]\right] \\ + \delta\left[\frac{M-1}{N-1}V_{M-1} + \left(1-\frac{M-1}{N-1}\right)V_M\right] \end{aligned}$$

We thus distinguish two cases.

(A) If $D_M(u) < \overline{s}$ for all $M \geq k$, then[100]

$$\begin{aligned} T_M\big(V(u),D(u)\big) &= \mathbb{E}\big[s\mathbf{1}\{\hat{s}-s \leq D_M\}\big] + \mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s > D_M\}\big] + \mathbb{P}(\hat{s}-s \leq D_M)D_M \\ &\quad + \delta\left[\frac{M-1}{N-1}V_{M-1} + \left(1-\frac{M-1}{N-1}\right)V_M\right] \\ &= h(D_M) + \delta\left[\frac{M-1}{N-1}V_{M-1} + \left(1-\frac{M-1}{N-1}\right)V_M\right] \end{aligned} \tag{30}$$

Consequently, if $D_M(u) < \overline{s}$ [101], plugging the above expressions in the equality $\Upsilon_M(u) = T_{M+1}(V,D) - T_M(V,D)$, and using the expression of D_M as a function of u, yields

$$\begin{aligned} \Upsilon_M(u) &= h(D_{M+1}) - h(D_M) + \delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ &= h(D_{M+1}) + h_1(D_M) - b \end{aligned} \tag{31}$$

[99] Indeed, for any joint distribution of types, the quantity

$$\mathbb{E}[(s+b)\mathbf{1}\{\hat{s}-s \leq X\}] + \mathbb{E}[\hat{s}\mathbf{1}\{\hat{s}-s > X\}]$$

decreases with $X \geq b$.

[100] Note that in this case the mapping T can be defined as $T : \big[0, V_{k-1} + k\overline{u}\big]^k \times \big[b, b+\overline{u}\big]^k \longrightarrow \big[0, V_{k-1} + k\overline{u}\big]^k$.

[101] By monotonicity (as $u \in K$), $D_M(u) < \overline{s}$ implies that $D_{M'} < \overline{s}$ for any $M' > M$.

Since $u \in K$, we have that (i) $u_M \geq 0$ for any M and thus by construction $D_M \geq b$, and (ii) the sequence $(u_M)_M$ is decreasing, and thus so is the sequence $(D_M)_M$. As a consequence, $D_M \geq D_{M+1} \geq b$.

Henceforth, we restrict our attention to joint distributions such that the functions h_1 and $(h-h_1)$ are strictly increasing over $[b,+\infty) \cap \text{Supp}(\hat{s}-s)$ [102]. This set notably includes the set of continuous joint symmetric distributions[103], as well as the case where the majority candidate has a fixed type $s \geq 0$ and the minority candidate a type $s+D$ where D is a (full support) random variable with a continuously differentiable distribution over $(-s,s)$ symmetric around 0.[104]

As a consequence, for any $u \in K$, $\Upsilon_M(u) \geq 0$ and the sequence $(\Upsilon_M(u))_{M \geq k}$ is decreasing as it inherits the monotonicity of the sequence $(D_M)_M$. Moreover, for any $M \geq k$,

$$\Upsilon_M(u) \leq h(D_M) + h_1(D_M) - b = \delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1 - \frac{M}{N-1}\right)u_M\right] < \delta\frac{N-2}{N-1}u_{k-1} \leq \overline{u}$$

It thus remains to check that $\Upsilon_{k-1}(u) \geq \Upsilon_k(u)$. By monotonicity of h and $(h-h_1)$ and using the above computations, a sufficient condition for this inequality to hold writes as:

$$(1-\delta)V_{k-1} \leq h(b)$$

This condition imposes an upper bound on V_{k-1}. Recall that $h(b)$ is the highest flow payoff (quality and homophily) that a majority member can guarantee. Therefore, for any symmetric joint distribution of types, any (increasing and concave) equilibrium value function must satisfy $V_{k-1} < h(b)/(1-\delta)$. Hence assuming this inequality hold does not threaten the existence of an equilibrium. We thus fix in the following V_{k-1} such that the above inequality holds. Hence, under the above conditions, $\Upsilon(u) \in K$.

(B) We now consider the case where $\overline{s} < +\infty$ and $D_M(u) = \overline{s}$ for some M. (Note that as $u_M \leq \overline{u} < \infty$, the case $D_M(u) = \overline{s}$ can only arise when $\overline{s} < \infty$.)

We first note that, within the class of equilibria with $u \in K$ (and thus a decreasing sequence $(\Delta_M)_M$), $\Delta_k = \overline{s}$ implies that $\Delta_{k+1} < \overline{s}$. Hence, whenever the majority is not tight, it recruits a minority candidate with a strictly positive probability: $\Delta_M < \overline{s}$ for any $M \geq k+1$.[105]

[102] Note that $(h-h_1)$ being strictly increasing implies that h is strictly increasing, as $h(X) - h_1(X) = 2h(X) - X$.

[103] Indeed, letting F be the marginal c.d.f. of s and $\hat{s}$, then

$$\forall \Delta > 0, \qquad h(\Delta) = \int_0^{\overline{s}} (s+\Delta)F(s+\Delta)dF(s) + \int_{\Delta}^{\overline{s}} \hat{s}F(\hat{s}-\Delta)dF(\hat{s}),$$

and thus, for any $\Delta \in (0,\overline{s})$,

$$h'(\Delta) = \int_0^{\overline{s}} F(s+\Delta)dF(s) + \int_0^{\overline{s}-\Delta} (s+\Delta)f(s+\Delta)dF(s) - \int_{\Delta}^{\overline{s}} \hat{s}f(\hat{s}-\Delta)dF(\hat{s}) = \int_0^{\overline{s}} F(s+\Delta)dF(s) \in (1/2,1)$$

since $\int_0^{\overline{s}} F(s)dF(s) = 1/2$.

[104] Indeed, denoting by F the c.d.f. of D, we have for any $\Delta \in (0,\overline{s})$,

$$h(\Delta) = \int_{-s}^{\Delta} (s+\Delta)dF(D) + \int_{\Delta}^{s} (s+D)dF(D), \qquad \text{and thus} \qquad h'(\Delta) = F(\Delta) \in (1/2,1)$$

[105] Indeed, suppose by contradiction that $\Delta_k = \Delta_{k+1} = \overline{s}$. Then, by construction,

$$u_k = \delta\Big[\frac{k}{N-1}u_k + \frac{k-2}{N-1}u_{k+1}\Big]$$

Since $u \in K$, this yields that $u_k = u_{k+1} = 0$, which contradicts the initial assumption as $b < \overline{s}$.

Consequently, we only need to consider the case where $D_{k+1}(u) < D_k(u) = \overline{s} < \infty$ [106]. We first show that $\Upsilon_k(u) \in [\Upsilon_{k+1}(u), \overline{u}]$. By construction,

$$T_k\big(V(u), D(u)\big) = \mathbb{E}[s] + b + \delta\left[\frac{k}{N-1}V_k + \left(1 - \frac{k}{N-1}\right)V_{k+1}\right],$$

and thus, since $D_{k+1} < \overline{s}$ implies that $T_{k+1}(V, D)$ is given by (30),

$$\Upsilon_k(u) = h(D_{k+1}) - \mathbb{E}[s] - b$$

By monotonicity of the sequence $(D_M)_M$ and since the functions h and h_1 are increasing, we have that $\Upsilon_k(u) \geq \Upsilon_{k+1}(u)$. It thus remains to check that $\Upsilon_{k-1}(u) \geq \Upsilon_k(u)$. A sufficient condition for this inequality to hold writes as[107]

$$(1-\delta)V_{k-1} \leq \mathbb{E}[s] + b + \frac{k}{N-2}(\overline{s} - b)$$

This second inequality is looser than the condition[108] in case (A) and is thus satisfied for $V_{k-1} \leq h(b)/(1-\delta)$ (which must be the case in any equilibrium as discussed above).

Therefore, fixing $V_{k-1} \in [0, h(b)/(1-\delta)]$, Υ is a well-defined continuous mapping from K into itself. By Brouwer's fixed point theorem, it admits a fixed point. This establishes existence.

We prove that in any equilibrium in this class, the sequence $(\Delta_M)_M$ is *strictly* decreasing. For the sake of exposition, we focus on the case $\Delta_k < \overline{s}$ (the case $\Delta_k = \overline{s}$ relies on a similar – and shorter – argument[109]). We suppose by contradiction that for some M, $\Delta_M = \Delta_{M+1}$ and note with (31) that this

[106] Indeed, note that if $D_{k+1}(u) < \overline{s}$, then $D_{k+1}(\Upsilon(u)) < \overline{s}$ as

$$\begin{aligned} D_{k+1}(\Upsilon(u)) &< b + \delta\Big[\frac{k}{N-1}\Big(h(D_{k+1}(u)) - \mathbb{E}[s] - b\Big) + \frac{k-2}{N-1}\Big(h(D_{k+2}(u) + h_1(D_{k+1}(u)) - b\Big)\Big] \\ &< \Big(1 - \delta\frac{N-2}{N-1}\Big)b + \delta\Big[\frac{k}{N-1}\big(h(D_{k+1}(u)) - \mathbb{E}[s]\big) + \frac{k-2}{N-1}D_{k+1}(u)\Big] \\ &< \Big(1 - \delta\frac{N-2}{N-1}\Big)b + \delta\frac{N-2}{N-1}\overline{s} < \overline{s} \end{aligned}$$

[107] Indeed, a sufficient condition for $\Upsilon_{k-1}(u) \geq \Upsilon_k(u)$ is

$$2(\mathbb{E}[s] + b) - (1-\delta)V_{k-1} + \delta u_{k-1} \geq h\Big(b + \delta\frac{N-2}{N-1}u_k\Big) - \delta\frac{k-1}{N-1}u_k,$$

which by monotonicity of h and $h - h_1$ holds in particular if

$$\begin{aligned} &2(\mathbb{E}[s] + b) - (1-\delta)V_{k-1} + \delta u_{k-1} \geq h\Big(b + \delta\frac{N-2}{N-1}u_{k-1}\Big) - \delta\frac{k-1}{N-1}u_{k-1}, \\ \text{i.e.} \quad &(1-\delta)V_{k-1} \leq 2(\mathbb{E}[s] + b) - h\Big(b + \delta\frac{N-2}{N-1}u_{k-1}\Big) + \delta\Big(1 + \frac{k-1}{N-1}\Big)u_{k-1} \end{aligned}$$

Hence, by monotonicity of $X \mapsto X - h(X)$ and since u_{k-1} must satisfy $\delta(N-2)/(N-1)u_{k-1} \geq (\overline{s} - b)$, a sufficient condition for this inequality to hold is

$$(1-\delta)V_{k-1} \leq 2(\mathbb{E}[s] + b) - h(\overline{s}) + (\overline{s} - b) + \frac{k}{N-2}(\overline{s} - b),$$

which yields the result as $h(\overline{s}) = \mathbb{E}[s] + \overline{s}$.

[108] Indeed, for any joint distribution such that $(\hat{s} - s)$ is symmetrically distributed around 0,

$$h(b) \leq \mathbb{E}[s] + b + \frac{k}{N-2}(\overline{s} - b)$$

[109] Indeed, we know from before that $\Delta_k = \overline{s}$ implies $\Delta_{k+1} < \Delta_k$.

implies $\Delta_M = \Delta_{M'}$ for any M, M' [110], and consequently $u_M = u_{M'}$ for any M, M'. Hence (31) implies that

$$u_M = h(\Delta_M) + h_1(\Delta_M) - b = \Delta_M - b = \delta \frac{N-2}{N-1} u_M,$$

where the last equality follows by definition of Δ_M, and thus $u_M = 0$ and $\Delta_M = b$ for all $M \geq k$. Therefore, $u_{k-1} = 0$. Hence, considering the minority's value function yields similarly that $u_i = 0$ for all $i \leq k-1$ [111]. However, by summing the expression of u_i for all $i \in \{1, ..., N-2\}$ and rearranging, yields on the LHS a weighted sum of u_i, which is thus equal to 0, while on the RHS the term $b\big[\mathbb{P}(\hat{s} - s \leq b) - \mathbb{P}(\hat{s} - s > b)\big]$ [112], which is strictly positive as $(\hat{s} - s)$ is symmetrically distributed around 0. Hence we reach a contradiction, and therefore, the sequence $(\Delta_M)_{M \geq k}$ is strictly decreasing.

We now turn to ranking equilibria from more to less meritocratic. Consider the class of equilibria characterized by a decreasing decision rule $(\Delta_M)_{M \in \{k,...,N-1\}}$. We refer in the following to an equilibrium by its decision rule $\Delta \equiv (\Delta_M)_{M \in \{k,...,N-1\}}$. Let Δ and Δ' be two equilibria within this class. We now show that

(i) $\Delta_k < \Delta'_k$ implies that $\Delta_M < \Delta'_M$ for any $M \geq k+1$,

(ii) $\Delta_k = \Delta'_k \in [b, \overline{s}]$ implies that $\Delta_M = \Delta'_M < \overline{s}$ for any $M \geq k+1$,

(i) Assume that $\Delta_k < \Delta'_k < \overline{s}$ (computations are analogous in the case $\Delta_k < \Delta'_k = \overline{s}$). By monotonicity, $\Delta_M < \overline{s}$ and $\Delta'_M < \overline{s}$ for any $M \geq k+1$, and thus, with the above notation,

$$\begin{aligned}\Delta_M &= b + \delta\left[\frac{M-1}{N-1} u_{M-1} + \left(1 - \frac{M}{N-1}\right) u_M\right] \\ &= \left(1 - \delta\frac{N-2}{N-1}\right) b + \delta\left[\frac{M-1}{N-1}\big[h(\Delta_M) + h_1(\Delta_{M-1})\big] + \left(1 - \frac{M}{N-1}\right)\big[h(\Delta_{M+1}) + h_1(\Delta_M)\big]\right]\end{aligned}$$

Consequently, for any $M \geq k+1$,

$$h_{2,M}(\Delta_M) - h_{2,M}(\Delta'_M) = \delta\frac{M-1}{N-1}\left[h_1(\Delta_{M-1}) - h_1(\Delta'_{M-1})\right] + \delta\left(1 - \frac{M}{N-1}\right)\left[h(\Delta_{M+1}) - h(\Delta'_{M+1})\right] \tag{32}$$

[110] Indeed, using the expression of Δ_M (for $\Delta_M < \overline{s}$), $\Delta_M = \Delta_{M+1}$ first gives that

$$\frac{M-1}{N-1}(u_M - u_{M-1}) + \left(1 - \frac{M+1}{N-1}\right)(u_{M+1} - u_M) = 0,$$

and thus, since the sequence $(u_M)_M$ is decreasing, $u_{M-1} = u_M = u_{M+1}$. The expression of u as a function of Δ (i.e. (31)), together with the strict monotonicity of the functions h and h_1 then implies that $\Delta_{M-1} = \Delta_M = \Delta_{M+1} = \Delta_{M+2}$.

[111] Indeed, when $\Delta_M = b$ for all $M \geq k$, then for any $i \leq k-2$,

$$u_i = \delta\mathbb{P}(\hat{s} - s \leq b)\left[\frac{i-1}{N-1} u_{i-1} + \left(1 - \frac{i}{N-1}\right) u_i\right] + \delta\mathbb{P}(\hat{s} - s > b)\left[\frac{i}{N-1} u_i + \left(1 - \frac{i+1}{N-1}\right) u_{i+1}\right],$$

The result follows by induction.

[112] Indeed, computations yield that

$$u_{k-1} = b\big[\mathbb{P}(\hat{s} - s \leq b) - \mathbb{P}(\hat{s} - s > b)\big] + \delta\mathbb{P}(\hat{s} - s \leq b)\left[\frac{k-2}{N-1} u_{k-2} + u_{k-1} + \frac{k-1}{N-1} u_k\right]$$

where the function $h_{2,M}$ is given by

$$h_{2,M}(X) \equiv X - \delta\frac{M-1}{N-1}h(X) - \delta\left(1-\frac{M}{N-1}\right)h_1(X),$$

We note that $h_{2,M}$ is strictly increasing over $[b,\overline{s}]$ [113]. By monotonicity of h_1, we get for $M = k+1$ that

$$h_{2,k+1}(\Delta_{k+1}) - h_{2,k+1}(\Delta'_{k+1}) < \delta\left(1-\frac{k+1}{N-1}\right)\left[h(\Delta_{k+2}) - h(\Delta'_{k+2})\right]$$

Suppose by contradiction that $\Delta_{k+1} \geq \Delta'_{k+1}$. Then by monotonicity, $\Delta_{k+2} \geq \Delta'_{k+2}$. By summing Equation (32) in $k+1$ and $k+2$ and rearranging, we get that

$$\begin{aligned}
&\left[h_{2,k+1}(\Delta_{k+1}) - \delta\frac{k+1}{N-1}h_1(\Delta_{k+1})\right] - \left[h_{2,k+1}(\Delta'_{k+1}) - \delta\frac{k+1}{N-1}h_1(\Delta'_{k+1})\right] \\
&\qquad + \left[h_{2,k+2}(\Delta_{k+2}) - \delta\frac{k-2}{N-1}h(\Delta_{k+2})\right] - \left[h_{2,k+2}(\Delta'_{k+2}) - \delta\frac{k-2}{N-1}h(\Delta'_{k+2})\right] \\
&= \delta\frac{k}{N-1}\left[h_1(\Delta_k) - h_1(\Delta'_k)\right] + \delta\left(1-\frac{k+2}{N-1}\right)\left[h(\Delta_{k+3}) - h(\Delta'_{k+3})\right]
\end{aligned}$$

Since for any $M \geq k+1$, the functions $h_{2,M} - \delta\frac{M}{N-1}h_1$ and $h_{2,M} - \delta\frac{N-M}{N-1}h$ are strictly increasing over $[b,\overline{s}]$, the above equality implies that $\Delta_{k+3} \geq \Delta'_{k+3}$. We now proceed by induction: suppose that $\Delta_j \geq \Delta'_j$ for any $j \in \{k+1,...,M\}$. Then by summing Equation (32) over the indices $k+1,...,M$ and rearranging,

$$\begin{aligned}
&\left[h_{2,k+1}(\Delta_{k+1}) - \delta\frac{k}{N-1}h_1(\Delta_{k+1})\right] - \left[h_{2,k+1}(\Delta'_{k+1}) - \delta\frac{k}{N-1}h_1(\Delta'_{k+1})\right] \\
&\quad + \left[h_{2,M}(\Delta_M) - \delta\frac{N-M}{N-1}h(\Delta_M)\right] - \left[h_{2,M}(\Delta'_M) - \delta\frac{N-M}{N-1}h(\Delta'_M)\right] \\
&\quad + \sum_{j=k+2}^{M-1}\left(\left[h_{2,j}(\Delta_j) - \delta\frac{j}{N-1}h_1(\Delta_j) - \delta\frac{N-j}{N-1}h(\Delta_j)\right] - \left[h_{2,j}(\Delta'_j) - \delta\frac{j}{N-1}h_1(\Delta'_j) - \delta\frac{N-j}{N-1}h(\Delta'_j)\right]\right) \\
&= \delta\frac{k-1}{N-1}\left[h_1(\Delta_k) - h_1(\Delta'_k)\right] + \delta\left(1-\frac{M}{N-1}\right)\left[h(\Delta_{M+1}) - h(\Delta'_{M+1})\right]
\end{aligned}$$

Since for any $j \geq k+1$, the functions $h_{2,j} - \delta\frac{j}{N-1}h_1 - \delta\frac{N-j}{N-1}h$ are strictly increasing over $[b,\overline{s}]$, we get that $\Delta_{M+1} \geq \Delta'_{M+1}$. Hence by induction, we have that $\Delta_M \geq \Delta'_M$ for any $M \geq k+1$. But by summing (32) over all these indices and rearranging yields

$$\begin{aligned}
0 \leq\ &\left[h_{2,k+1}(\Delta_{k+1}) - \delta\frac{k}{N-1}h_1(\Delta_{k+1})\right] - \left[h_{2,k+1}(\Delta'_{k+1}) - \delta\frac{k}{N-1}h_1(\Delta'_{k+1})\right] \\
&+ \left[h_{2,N-1}(\Delta_{N-1}) - \delta\frac{1}{N-1}h(\Delta_{N-1})\right] - \left[h_{2,N-1}(\Delta'_{N-1}) - \delta\frac{1}{N-1}h(\Delta'_{N-1})\right] \\
&+ \sum_{j=k+2}^{N-2}\left(\left[h_{2,M}(\Delta_M) - \delta\frac{M}{N-1}h_1(\Delta_M) - \delta\frac{N-M}{N-1}h(\Delta_M)\right] - \left[h_{2,M}(\Delta'_M) - \delta\frac{M}{N-1}h_1(\Delta'_M) - \delta\frac{N-M}{N-1}h(\Delta'_M)\right]\right) \\
=\ &\delta\frac{k-1}{N-1}\left[h_1(\Delta_k) - h_1(\Delta'_k)\right] < 0
\end{aligned}$$

[113] Indeed, we may rewrite the function $h_{2,M}$ as: $h_{2,M}(X) = \left[1-\delta\left(1-\frac{M}{N-1}\right)\right]h_1(X) + \left[1-\delta\frac{M-1}{N-1}\right]h(X)$.

which is a contradiction. Therefore, $\Delta_{k+1} < \Delta'_{k+1}$. The result then obtains by induction, supposing by contradiction that $\Delta_j < \Delta'_j$ for any $j \in \{k, ..., M-1\}$ and that $\Delta_M \geq \Delta'_M$, and considering the sums of (32) over appropriate indices so as to reach a contradiction.
(ii) We note that the above argument yields that if $\Delta_k = \Delta'_k \in [b, \overline{s}]$, then $\Delta_M = \Delta'_M$ for any $M \geq k+1$. As a consequence, any two distinct equilibria with a decreasing decision rule satisfy either "$\Delta_M < \Delta'_M$ for all $M \geq k$", or "$\Delta_M > \Delta'_M$ for all $M \geq k$".

Lastly, we turn to Pareto-comparing the equilibria. Consider two equilibria within this class, described by a decreasing decision rule denoted respectively by Δ and Δ' such that $\Delta \prec \Delta'$, and let $(V_i)_{i\in\{1,...,N-1\}}$ and $(V'_i)_{i\in\{1,...,N-1\}}$ be the corresponding equilibrium value functions. For any $M \geq k$, we have by construction that

$$V_M = \mathbb{E}[(s+b)\mathbf{1}\{\hat{s} - s \leq \Delta_M\}] + \mathbb{E}[\hat{s}\mathbf{1}\{\hat{s} - s > \Delta_M\}] + \delta\mathbb{P}(\hat{s} - s \leq \Delta_M)\left[\frac{M}{N-1}V_M + \left(1 - \frac{M}{N-1}\right)V_{M+1}\right]$$
$$+ \delta\big(1 - \mathbb{P}(\hat{s} - s \leq \Delta_M)\big)\left[\frac{M-1}{N-1}V_{M-1} + \left(1 - \frac{M-1}{N-1}\right)V_M\right]$$

Note that $\Delta_k < \Delta'_k$ implies that $\Delta_k < \overline{s}$. Hence, using that for any $M \geq k$, $\Delta_M = b + \delta\left[\frac{M-1}{N-1}u_{M-1} + \frac{N-M-1}{N-1}u_M\right]$, yields

$$\left[1 - \delta\left(1 - \frac{M-1}{N-1}\right)\big[1 - \mathbb{P}(\hat{s} - s \leq \Delta'_M)\big] - \delta\frac{M}{N-1}\mathbb{P}(\hat{s} - s \leq \Delta'_M)\right](V_M - V'_M) \tag{33}$$
$$= \mathbb{E}\big[(\hat{s} - s - \Delta_M)\mathbf{1}\{\Delta_M < \hat{s} - s \leq \Delta'_M\}\big] + \delta\mathbb{P}(\hat{s} - s \leq \Delta'_M)\left(1 - \frac{M}{N-1}\right)(V_{M+1} - V'_{M+1})$$
$$+ \delta\big(1 - \mathbb{P}(\hat{s} - s \leq \Delta'_M)\big)\frac{M-1}{N-1}(V_{M-1} - V'_{M-1})$$

Two cases arise depending on whether $\Delta'_k = \overline{s}$. If so, then the result for majority members follows by the usual argument (by contradiction and by induction). Hence, for any $\delta \in [0, (N-1)/N)$, any "meritocratic" equilibrium (i.e. with $\Delta_k < \overline{s}$) is preferred at any majority size by all majority members to the entrenched equilibrium ($\Delta'_k = \overline{s}$).

By contrast, if $\Delta'_k < \overline{s}$, one needs to investigate the minority value function as well. We thus focus on the case $\Delta'_k < \overline{s}$. We use the same argument as in the proof of Lemma 1 and Proposition 2. Suppose by contradiction that $V_{N-1} \leq V'_{N-1}$. Then equation (33) implies that $V_{N-2} - V'_{N-2} \leq V_{N-1} - V'_{N-1} \leq 0$, and thus by induction that $V_{k-1} - V'_{k-1} \leq V_k - V'_k \leq ... \leq V_{N-1} - V'_{N-1} \leq 0$.

Similarly, for any $i \leq k-1$, we have by construction that

$$V_i = \mathbb{E}[s\mathbf{1}\{\hat{s} - s \leq \Delta_{N-1-i}\}] + \mathbb{E}[(\hat{s}+b)\mathbf{1}\{\hat{s} - s > \Delta_{N-1-i}\}]$$
$$+ \delta\mathbb{P}(\hat{s} - s \leq \Delta_{N-1-i})\left[\frac{i-1}{N-1}V_{i-1} + \left(1 - \frac{i-1}{N-1}\right)V_i\right]$$
$$+ \delta\big(1 - \mathbb{P}(\hat{s} - s \leq \Delta_{N-1-i})\big)\left[\frac{i}{N-1}V_i + \left(1 - \frac{i}{N-1}\right)V_{i+1}\right]$$

Hence, for any $i \leq k-1$,

$$\left[1-\delta\left(1-\frac{i-1}{N-1}\right)\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})-\delta\frac{i}{N-1}\left[1-\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})\right]\right](V_i-V'_i) \tag{34}$$
$$=\mathbb{E}\left[\left[\hat{s}-s+b+\delta\left(\frac{i-1}{N-1}u_{i-1}+\frac{N-1-i}{N-1}u_i\right)\right]\mathbf{1}\{\Delta_{N-1-i}<\hat{s}-s\leq\Delta'_{N-1-i}\}\right]$$
$$+\delta\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})\frac{i-1}{N-1}(V_{i-1}-V'_{i-1})+\delta\big(1-\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})\big)\left(1-\frac{i}{N-1}\right)(V_{i+1}-V'_{i+1})$$

Hence, for δ close to 0, the expectation term on the RHS of (34) is strictly positive. Suppose by contradiction that $V_1 \leq V'_1$. Then, by induction, equation (34) yields that $V_k - V'_k \leq ... \leq V_1 - V'_1 \leq 0$. However, by summing (33) and (34) over all indices and rearranging yields on one side a (positively) weighted sum of differences $(V_i - V'_i)$, which is thus negative, while on the other side the sum of flow payoffs, i.e. the sum of the expectation terms on the RHS of (33) and (34), which is strictly positive. Hence a contradiction. Therefore, $V_1 > V'_1$.

Working in a similar fashion – by contradiction and by induction, reaching the desired contradiction by considering sums of (33) and (34) over appropriate indices – yields the result: for all $i \in \{1, ..., N-1\}$, $V_i > V'_i$.

H Proof of Proposition 5

The properties of the value functions of the two canonical equilibria with homogamic evaluation capability depend on whether $x^\dagger \leq 1/2$. If $x^\dagger \leq 1/2$, they exhibit the same features – monotonicity and concavity/convexity – as their perfect-information counterparts (indeed, the proof of Lemma 1 goes through replacing x by $x^\dagger$). By contrast, if $x^\dagger > 1/2$, the value function in the meritocratic equilibrium (if it exists) now decreases with group size $i \in \{1, ..., N-1\}$ [This observation immediately gives that for $x^\dagger > 1/2$, the meritocratic equilibrium exists for any $s^\dagger > b$.], and is concave for the minority ($i \leq k-1$) and convex for the majority ($i \geq k$). Similarly, in the entrenched equilibrium (if it exists), the value function increases less over $\{k, ..., N-1\}$ than it decreases over $\{1, ..., k-1\}$, whereas with $x^\dagger \leq 1/2$, the opposite holds: the distinction stems from the fact that the (weighted) sum of differences $V^{\mathrm{e}}_{i+1} - V^{\mathrm{e}}_i$ is equal to $(1-2x^\dagger)b$. As a consequence, with $x^\dagger \geq 1/2$, in the entrenchment equilibrium, it is not the case in general that $V^{\mathrm{e}}_i \geq V^{\mathrm{e}}_{N-i-1}$ for any $i \geq k$, while in the meritocratic equilibrium, $V^{\mathrm{m}}_i \leq V^{\mathrm{m}}_{N-i-1}$ for any $i \geq k$ (the curse of control in action).

Let the quantities $Y^\dagger$ and $Z^\dagger$ be given by

$$\begin{cases} Y^\dagger \equiv 1+\delta\dfrac{k-1}{N-1}x^\dagger\displaystyle\sum_{t=0}^{+\infty}\delta^t\left(\pi^{\mathrm{e}\dagger}_{k+1,k}(t)-\hat{\pi}^{\mathrm{e}\dagger}_{k,k}(t)\right) \\ Z^\dagger \equiv 1+\dfrac{k-1}{N-1}\dfrac{\delta}{1-\delta}(1-2x^\dagger)+\delta\dfrac{k-1}{N-1}x^\dagger\displaystyle\sum_{t=0}^{+\infty}\delta^t\left(\pi^{\mathrm{e}\dagger}_{k+1,k}(t)+\hat{\pi}^{\mathrm{e}\dagger}_{k,k}(t)\right) \end{cases}$$

where the probabilities $\pi^{\mathrm{e}\dagger}_{i,j}(t)$ (resp. $\hat{\pi}^{\mathrm{e}\dagger}_{i,j}(t)$) are taken *(a)* following the entrenched equilibrium strategies described in Proposition 5, and *(b)* from a majority member's perspective (resp. minority member's

perspective) with transition parameter $x^\dagger$ instead of x. Define then $\rho^{\mathrm{e}\dagger}$ as

$$\rho^{\mathrm{e}\dagger} \equiv \begin{cases} \dfrac{x^\dagger}{x}\dfrac{Z^\dagger}{Y^\dagger} & \text{if } Y^\dagger > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

The same argument as the one used in the proof of $\rho^{\mathrm{e}} < +\infty$ [114] yields that for any $\delta \in [0, (N-1)/N)$ and $x^\dagger \in [0,1)$, $\rho^{\mathrm{e}\dagger} < \infty$.

Similarly, let $\rho^{\mathrm{m}\dagger}$ be defined as

$$\rho^{\mathrm{m}\dagger} \equiv \frac{x^\dagger}{x}\Big[1 + \frac{k-1}{N-1}(1-2x^\dagger)\delta\sum_{t=0}^{+\infty}\delta^t\Big[\Big(\sum_{i=k}^{N-1}\pi^{\mathrm{m}\dagger}_{k+1,i}(t)\Big) - \Big(\sum_{i=k}^{N-1}\pi^{\mathrm{m}\dagger}_{k-1,i}(t)\Big)\Big]\Big]$$

where the probabilities $\pi^{\mathrm{m}\dagger}_{i,j}(t)$ are taken *(a)* following the meritocratic equilibrium strategies described in Proposition 5, and *(b)* from the perspective of a member of the group with initial size i, with transition parameter $x^\dagger$ instead of x. We show that the thresholds $\rho^{\mathrm{m}\dagger}$ and $\rho^{\mathrm{e}\dagger}$ are the homogamic-evaluation-capability counterparts of ρ^{m} and ρ^{e} in the baseline setting.

The proof of Proposition 5 is analogous to that of Proposition 1. As mentioned, when $x^\dagger \leq 1/2$, the value functions in the entrenched and meritocratic equilibria with homogamic evaluation capability exhibit features similar to the ones of their perfect-information counterparts. Namely, the sequence $(V^{\mathrm{e}\dagger}_M)_{M\geq k}$ remains increasing and concave. By contrast, the monotonicity of the sequence $(V^{\mathrm{m}\dagger}_M)_{M\geq k}$ may differ: it is increasing (and concave) if $x^\dagger \leq 1/2$, whereas it is decreasing (and convex) if $x^\dagger > 1/2$. Moreover, in this latter case it may then be that $V^{\mathrm{e}\dagger}_k < V^{\mathrm{e}\dagger}_{k-1}$. Nonetheless, for $x^\dagger > 1/2$, the sequence $(V^{\mathrm{m}\dagger}_M)_{M\geq k}$ being decreasing implies that its differences $(V^{\mathrm{m}\dagger}_{M+1} - V^{\mathrm{m}\dagger}_M)$ are negative and thus recruiting the minority candidate against an untalented majority candidate is optimal (as $s^\dagger > b$): hence, for $x^\dagger > 1/2$, the meritocratic equilibrium exists whenever $s^\dagger > b$. Lastly, in both cases, because of discounting, a talented majority candidate is still preferred to the minority candidate (with unknown talent) at any majority size.

We thus consider $x^\dagger \in [0,1]$ henceforth. As noted above, the argument used in step 1 of the proof of Proposition 1 applies to both equilibria[115], thus yielding that (except in the meritocratic equilibrium for $x^\dagger > 1/2$), the most profitable deviation from these candidate equilibria is when the majority is tight and faces an untalented majority candidate together with an unknown-quality minority one. We thus focus on step 2 and consider one-shot deviations in majority size $M = k$ when the majority candidate is untalented.

Note that the difference between the expected maximum of both candidates' talents and the expected quality of the majority candidate writes as before $(\overline{x} + (1-\overline{x})x/x^\dagger)s - \overline{x}s = xs$.

[114] Cf. Section C.3.

[115] For both equilibria when $x^\dagger \leq 1/2$ and for the entrenchment equilibrium when $x^\dagger \geq 1/2$, the argument goes through replacing x by $x^\dagger$ and s by $s^\dagger$ when appropriate. In particular, in the entrenched equilibrium, for $x^\dagger \in [0,1]$, analogous computations yield that

$$\delta\Big(\frac{k-2}{N-1}u^{\mathrm{e}}_{k+1} + \frac{k}{N-1}u^{\mathrm{e}}_k\Big) \leq \frac{\delta\dfrac{k}{N-1}}{1-\delta\dfrac{k}{N-1}}\frac{1}{1-\overline{x}}\big(xs - (1-\overline{x})b\big) < s^\dagger - b$$

Hence a (one-shot) deviation in majority size k from the entrenched strategy (defined in Proposition 5), i.e. picking the minority candidate (of unknown talent) instead of the untalented majority candidate, yields a payoff equal to:

$$\Delta^{\mathrm{e},\dagger} \equiv s^\dagger - b + \delta\frac{k-1}{N-1}xs\sum_{t=0}^{+\infty}\delta^t\left(\pi^{\mathrm{e}\dagger}_{k+1,k}(t) - \hat{\pi}^{\mathrm{e}\dagger}_{k,k}(t)\right) + \delta\frac{k-1}{N-1}x^\dagger b\sum_{t=0}^{+\infty}\delta^t\left(\sum_{i\geq k+1}\hat{\pi}^{\mathrm{e}\dagger}_{k,i}(t)\right)$$
$$- \delta\frac{k-1}{N-1}x^\dagger b\sum_{t=0}^{+\infty}\delta^t\pi^{\mathrm{e}\dagger}_{k+1,k}(t) - \frac{k-1}{N-1}\frac{\delta}{1-\delta}(1-x^\dagger)b$$

where the probabilities $\pi^{\mathrm{e}\dagger}_{i,j}(t)$ (resp. $\hat{\pi}^{\mathrm{e}\dagger}_{i,j}(t)$) are taken *(a)* following the entrenched equilibrium strategies described in Proposition 5, and *(b)* from a majority member's perspective (resp. minority member's perspective) with transition parameter $x^\dagger$ instead of x. By construction, $s^\dagger/s = x/x^\dagger$. Rearranging yields

$$\Delta^{\mathrm{e},\dagger} = \frac{x}{x^\dagger}s\left[1 + \delta\frac{k-1}{N-1}x^\dagger\sum_{t=0}^{+\infty}\delta^t\left(\pi^{\mathrm{e}\dagger}_{k+1,k}(t) - \hat{\pi}^{\mathrm{e}\dagger}_{k,k}(t)\right)\right]$$
$$- b\left[1 + \frac{k-1}{N-1}\frac{\delta}{1-\delta}(1-2x^\dagger) + \delta\frac{k-1}{N-1}x^\dagger\sum_{t=0}^{+\infty}\delta^t\left(\pi^{\mathrm{e}\dagger}_{k+1,k}(t) + \hat{\pi}^{\mathrm{e}\dagger}_{k,k}(t)\right)\right]$$

which yields the result for the existence region of the entrenched equilibrium.

Similarly for the meritocratic equilibrium, consider the (one-shot) deviation of a majority member voting in k the untalented majority candidate instead of the minority one. Such a deviation yields a payoff equal to:

$$\Delta^{\mathrm{m},\dagger} = b - s^\dagger + \delta\frac{(k-1)}{N-1}(1-x^\dagger)b\sum_{t=0}^{+\infty}\delta^t\left[\left(\sum_{i\geq k}\pi^{\mathrm{m}\dagger}_{k+1,i}(t)\right) - \left(\sum_{i\geq k}\pi^{\mathrm{m}\dagger}_{k-1,i}(t)\right)\right]$$
$$+ \delta\frac{(k-1)}{N-1}x^\dagger b\sum_{t=0}^{+\infty}\delta^t\left[\left(\sum_{i\leq k-1}\pi^{\mathrm{m}\dagger}_{k+1,i}(t)\right) - \left(\sum_{i\leq k-1}\pi^{\mathrm{m}\dagger}_{k-1,i}(t)\right)\right]$$

i.e. by rearranging,

$$\Delta^{\mathrm{m},\dagger} = -\frac{x}{x^\dagger}s + b\left[1 + \delta(1-2x^\dagger)\frac{(k-1)}{N-1}\sum_{t=0}^{+\infty}\delta^t\left[\left(\sum_{i\geq k}\pi^{\mathrm{m}\dagger}_{k+1,i}(t)\right) - \left(\sum_{i\geq k}\pi^{\mathrm{m}\dagger}_{k-1,i}(t)\right)\right]\right]$$

The result for the existence region of the meritocratic equilibrium follows. Lastly, the proof for $\rho^{\mathrm{e},\dagger} < +\infty$ is in Section C.3.

Note moreover that Lemma 4 holds with the transition probabilities $\pi^{\mathrm{e}\dagger}$ and $\pi^{\mathrm{m}\dagger}$ [116], and this establishes the inequality $\rho^{\mathrm{m}\dagger} < \rho^{\mathrm{e}\dagger}$ for $x^\dagger \leq 1/2$, as well as the inequality $\rho^{\mathrm{m}\dagger} \leq x^\dagger/x$ for $x^\dagger \geq 1/2$ (noted in the text).[117]

[116] Indeed, the proof holds for any $x \in [0,1]$ as the stochastic matrices P and $\hat{P}$ (introduced in the proof of Lemma 4) remain stochastically monotone and stochastically comparable (with P stochastically dominating $\hat{P}$) for any $x \in [0,1]$.

[117] If $b < s^\dagger$ and $x^\dagger \geq 1/2$, then $\rho^{\mathrm{m}\dagger} \leq x^\dagger/x$, and thus the meritocratic equilibrium exists for all $s/b \geq x^\dagger/x$. Lastly, $s^\dagger$ and $x^\dagger$ both depend on x, and thus the value of $x^\dagger$ constrains the possible values of $s^\dagger$: in particular, for $x^\dagger \geq 1/2$ (and thus $\alpha \leq 1/2$), $s^\dagger$ decreases with $x^\dagger$, and $s^\dagger = 0$ when $x^\dagger = 1$. As a consequence, for any $b > 0$, the inequality $s^\dagger > b$ can

I Proof of Proposition 6

For the sake of exposition, we first focus on the case $l = k-2$, before turning to the general proof for $l \in \{1, ..., k-2\}$, which derives from the same argument.

We first show existence. The argument is analogous to the one used in the proof of Proposition 11. Put succinctly, we show that for $s = b(> 0)$ the strategies of Proposition 6 describe a symmetric MPE in weakly undominated strategies. Since for $s = b$, any deviation from the equilibrium strategy is strictly unprofitable and since the deviation differential payoffs are continuous with respect to s, the result then obtains for s/b in a neighbourhood of 1. Next we claim that this equilibrium is in fact the unique *monotone* equilibrium for s/b sufficiently close to 1.

We define as before $u_i \equiv V_{i+1} - V_i$ for any $i \in \{1, ..., N-2\}$. As emphasized in the text, the majority controls the outcome of the vote for majority sizes $N-2$ and $N-1$, and thus the usual argument yields that meritocratic decisions are optimal from the majority's perspective in $N-1$ for any $s \geq b$. Moreover, with the strategies described in Proposition 6,

$$\left[1 - \delta(1-x)\frac{N-2}{N-1}\right]u_{N-2} = x(s-b) \geq 0,$$

and thus in particular $u_{N-2} = 0$ when $s = b$. Hence we thereafter focus on deviations for majority and minority members in majority sizes strictly below $N-1$. Explicit computations then give that with the strategies in Proposition 6, for any $M \in \{k, ..., N-3\}$:

$$\begin{aligned}\left[1 - \delta\Lambda(M)\left(1 - \frac{M}{N-1}\right) - \delta(1-\Lambda(M+1))\frac{M}{N-1}\right]u_M \\ = [\Lambda(M) - \Lambda(M+1)]b + \delta\Lambda(M)\frac{M-1}{N-1}u_{M-1} + \delta(1-\Lambda(M+1))\left(1 - \frac{M+1}{N-1}\right)u_{M+1},\end{aligned} \tag{35}$$

and

$$\begin{aligned}\left[1 - \delta\Lambda(M)\left(1 - \frac{M+1}{N-1}\right) - \delta(1-\Lambda(M+1))\frac{M+1}{N-1}\right]u_{N-M-2} \\ = [\Lambda(M) - \Lambda(M+1)]b + \delta\Lambda(M)\frac{M}{N-1}u_{N-M-1} + \delta(1-\Lambda(M+1))\left(1 - \frac{M+2}{N-1}\right)u_{N-M-3},\end{aligned} \tag{36}$$

while for $i = k-1$,

$$[1 - \delta(1-\Lambda(k))]u_{k-1} = (1-2\Lambda(k))b + \delta(1-\Lambda(k))\frac{k-1}{N-1}u_k + \delta(1-\Lambda(k))\frac{k-2}{N-1}u_{k-2} \tag{37}$$

The proof derives from the following observation: for $s = b$, the one-shot deviation differential payoff from the strategies in Proposition 6 is a weighted sum of two consecutives terms $V_i - V_{i+1} = -u_i$ for some $i \in \{1, ..., N-3\}$ [118]. Hence in order to reach the desired result, we show that for any $i \in \{1, ..., N-3\}$, $u_i > 0$.

only hold for $x^\dagger$ sufficiently below 1.

[118] Namely, for $s = b$, the one-shot deviation differential payoff at group size i writes as

$$-\delta\left(1 - \frac{i}{N-1}\right)u_i - \delta\frac{i-1}{N-1}u_{i-1}$$

We proceed by contradiction, adapting the argument used in the proof of Lemma 1. Suppose by contradiction that $u_{N-3} \leq 0$. Then, Equation (35) implies that $u_i \leq 0$ for any $i \in \{k, ..., N-4\}$. Indeed, this can be shown by induction: $u_{N-3} \leq 0$ and Equation (35) in $M = N-3$ give that $u_{N-4} \leq u_{N-3} \leq 0$. Equation (35) for $M = N-4$ then implies that

$$0 \geq \left[1 - \delta\Lambda(N-4)\left(1 - \frac{N-4}{N-1}\right) - \delta\frac{N-4}{N-1}\right]u_{N-4} \geq [\Lambda(N-4) - \Lambda(N-5)]b + \delta\Lambda(N-4)\frac{N-5}{N-1}u_{N-5}$$

and thus $u_{N-5} \leq u_{N-4} \leq 0$. The result thus obtains by downward induction on the majority size (supposing the result holds for any majority size strictly above M, supposing by contradiction it does not for M, and summing (35) for all the majority sizes strictly above M in order to reach a contradiction).

Similarly, if $u_1 \leq 0$, then the same induction argument implies that $u_i \leq 0$ for any $i \in \{2, ..., k-1\}$. However, by summing (35)-(36)-(37) on all indexes, one obtains that

$$0 \geq (1-\delta)\sum_{i=1}^{N-3} u_i = b + \frac{\delta}{N-1}u_{N-2} > 0,$$

which is a contradiction. Hence $u_1 > 0$. If $u_2 \leq 0$, then by the same argument as above: $u_i \leq 0$ for any $i \in \{2, ..., k-1\}$. Yet by summmation again,

$$\begin{aligned} 0 &\geq \left[1 - \delta + \delta\Lambda(N-3)\frac{N-3}{N-1}\right]u_2 + (1-\delta)\sum_{i=3}^{N-3} u_i \\ &= (1 - \Lambda(N-3))b + \frac{\delta}{N-1}u_{N-2} + \delta\frac{1-\Lambda(N-3)}{N-1}u_1 > 0, \end{aligned}$$

which is again a contradiction, and thus $u_2 > 0$. Hence, by induction (supposing the result holds for any minority size strictly below i, supposing by contradiction it does not for i, and summing (35) for all the minority sizes weakly above i and all majority sizes up to $N-3$ in order to reach a contradiction), $u_i > 0$ for any $i \leq k-2$. Yet then, by summation of (35) over indices $k-1$ to $N-3$,

$$\begin{aligned} 0 &\geq \left[1 - \delta + \delta\Lambda(k)\frac{k}{N-1}\right]u_{k-1} + (1-\delta)\sum_{i=k}^{N-3} u_i \\ &= (1 - \Lambda(k) - \Lambda(N-2))b + \delta(1-\Lambda(k))\frac{k-2}{N-1}u_{k-2} + \frac{\delta}{N-1}u_{N-2} > 0, \end{aligned}$$

which is a contradiction and therefore, $u_{N-3} > 0$. The argument is then repeated by assuming by contradiction that $u_{N-4} \leq 0$ and considering the appropriate summations and the same induction arguments in order to reach a contradiction.

Consequently, when $s = b$, $u_i > 0$ for $i \in \{1, ..., N-3\}$ and hence the strategies in Proposition 6 describe a symmetric MPE in weakly undominated strategies. Since all deviation differential payoffs are strictly negative, the argument extends by continuity to s in an upper neighbourhood of b. This establishes the existence of the equilibrium.

General proof of existence. Let $s = b > 0$. Consider any $l \in \{1, ..., k-2\}$ and the strategy of super-entrenchment to level l, denoting by V_i the corresponding value function and u_i its first-difference.

Since $s = b$, the usual computations[119] (see proof of Lemma 1) yield that for any $i \geq k + l$ and for any $i \leq k-2-l$, $u_i = 0$. The above computations then apply, using (35) for group sizes $i \in \{k, ..., k+l-1\}$, (36) for group sizes $i \in \{k-2-l, ..., k-2\}$, and (37) for group size $k-1$. The result obtains by continuity for s/b in a neighbourhood of 1.

Proof of uniqueness. We now show that, for s/b close to 1, super-entrenchment at level l is the unique symmetric MPE such that a stronger majority makes more meritocratic recruitments. Hence, we consider the class of equilibria such that a stronger majority makes more meritocratic recruitments, and show that, for any candidate equilibrium within this class, for $s = b > 0$, the majority is super-entrenched in $k + l$. By monotonicity, this implies that all candidate equilibria within this class must feature an entrenched majority at majority sizes $M \in \{k, ..., k + l\}$. We then show that the minority best-replies to this strategy by voting for the in-group candidate whenever it may be pivotal, i.e. at any majority size $M \leq k + l - 1$.

We begin by noting that when $s = b$, a group's flow payoff whenever it is pivotal does not depend on its being meritocratic or entrenched (as the difference between the two writes as $x(s - b) = 0$). Moreover, for $s = b$, the flow differential payoff in the expression of u_i writes as $[\Lambda(i) - \Lambda(i + 1)]b$ (resp. $[\Lambda(i)-\Lambda(i+1)](1-2x)b$) if the minority follows entrenchment (resp. meritocracy) at majority sizes i and $i+1$, as $[\Lambda(i)-\Lambda(i+1)]b-2x\Lambda(i)b$ if the minority follows meritocracy at majority size i and entrenchment at majority size $i + 1$, and as $[\Lambda(i) - \Lambda(i + 1)]b + 2x\Lambda(i + 1)b$ if the minority follows entrenchment at majority size i and meritocracy at majority size $i+1$. In particular, the flow-payoff term in u_{k+l-1} writes as $\Lambda(k + l - 1)b$ if the minority is entrenched at majority size $k + l - 1$ (resp. $\Lambda(k + l - 1)(1 - 2x)b$ if it votes meritocratically). By contrast, for any $i \geq k + l$, the flow payoff term in u_i is equal to 0.

We now show that the majority is always entrenched in $k + l$, i.e. that in any equilibrium,

$$\frac{k+l-1}{N-1}u_{k+l-1} + \left(1 - \frac{k+l}{N-1}\right)u_{k+l} > 0$$

Suppose by contradiction that the majority votes meritocratically at size $k + l$ (i.e. that the above LHS is weakly lower than 0). Suppose first that $u_{k+l} \leq 0$. Hence, the recursive expression of u_i for $i \geq k + l$ yields that $u_{k+l-1} \leq u_{k+l} \leq ... \leq u_{N-2} \leq 0$. Then, the recursive expression of u_{k+l-1} together with the above remark (positivity of the flow-payoff term) implies that $u_{k+l-2} < u_{k+l-1} \leq 0$. We proceed by induction. Our above remark on flow-payoff terms gives that "$u_{M+1} < u_{M+2} \leq 0$" implies "$u_M < u_{M+1} \leq 0$" as long as the minority does not change its strategy between sizes M and $M + 1$, or goes from entrenchment to meritocracy when going from majority size M to $M + 1$. The only case we need to discuss is the one where the minority goes from meritocracy to entrenchment when going from majority size M to $M+1$, i.e. intuitively when the minority tries to deter the majority from growing too strong. Summing (and rearranging as usual) the recursive expressions of the differential value function u_i over indices $i \in \{M, ..., N - 2\}$ then gives on the LHS a weighted sum of u_i for $i \in \{M, ..., N - 2\}$, which is weakly negative with the induction hypothesis, while on the RHS a first term proportional to u_{M-1} and a second term which is the sum of the flow-differential payoffs, equal to $\Lambda(M)(1 - 2x)b > 0$. Therefore, $u_{M-1} < u_M \leq 0$ [120].

[119] This could be seen by using the recursive expressions for the sequence $(u_i)_i$ and supposing by contradiction that $u_i \neq 0$ for some $i \geq k + l$ or $\leq k - 2 - l$.

[120] Indeed, $u_{M-1} \leq 0$ together with $u_M \leq 0$ and $u_{M+1} \leq 0$ implies that the majority must vote meritocratically at sizes

Hence, by induction, $u_i \leq 0$ for any $i \in \{k-1, ..., N-2\}$, and consequently, the majority is meritocratic at any majority size $i \geq k$. Therefore, the flow differential payoffs in the expression of u_i for $i \leq k-1$ write as $[\Lambda(i) - \Lambda(i+1)](1-2x)b > 0$ for any $i \in \{k-l-1, ..., k-2\}$, and 0 for any $i \leq k-l-2$.

Suppose by contradiction that $u_{k-l-1} \leq 0$. Then the recursive expression of u_i for $i \leq k-l-2$ yield that $u_{k-l-1} \leq ... \leq u_1 \leq 0$. Furthermore, since the flow differential payoffs are positive for $i \in \{k-l-1, ..., k-2\}$, we have that $u_i \leq 0$ for $i \in \{1, ..., k-1\}$. Therefore the minority votes meritocratically whenever it is pivotal. Hence, the sum of the flow differential payoffs over all indices $i \in \{1, ..., N-2\}$ writes as

$$2\Lambda(k)(1-2x)b + [1 - 2\Lambda(k)](1-2x)b = (1-2x)b > 0$$

where the second term is the flow differential payoff in u_{k-1}. Yet this contradicts $u_i \leq 0$ for all $i \in \{1, ..., N-2\}$.

Hence $u_{k-l-1} > 0$. The recursive expressions of the differential value function now yield $0 < u_1 < ... < u_{k-l-1}$. Supposing by contradiction that $u_{k-l} \leq 0$ yields again that $u_i \leq 0$ for $i \in \{k-l, ..., k-1\}$. Hence by summing the recursive expressions of u_i for $i \in \{k-l, ..., N-2\}$ and rearranging yields on the LHS a weighted sum of the differential value function u_i for $i \in \{k-l, ..., N-2\}$, which is weakly negative, while on the RHS, a term proportional to u_{k-l-1} (and thus strictly positive) and the sum of the flow differential payoffs, which is strictly positive. This is a contradiction, and thus $u_{k-l} > 0$. Using repeatedly the same argument, we have by induction that $u_i > 0$ for any $i \leq k-2$, and as a consequence, the minority is entrenched whenever it has size $i \in \{k-l, ..., k-2\}$, i.e. whenever the majority has size $i \in \{k+1, ..., k+l-1\}$. Summing again the recursive expression of the differential value function u_i over indices $i \geq k-1$ yields after rearranging, on the LHS a weighted sum of the differential value function u_i for $i \in \{k-1, ..., N-2\}$, which is weakly negative, while on the RHS, a term proportional to u_{k-2} (and thus strictly positive) and the sum of the flow differential payoffs, which is equal to $[1-\Lambda(k)](1-2x) > 0$. Hence the RHS is strictly positive, which is a contradiction. Therefore, $u_{k+l} > 0$, and thus using the recursive expression of u_i for $i \geq k+l$, we have that $u_{k+l-1} > u_{k+l} > u_{k+l+1} > ... > u_{N-2} > 0$ (as we suppose that the majority votes meritocratically at size $k+l$). This establishes that the majority would strictly benefit by deviating to entrenchment when it has size $k+l$, and thus contradicts the assumption that the majority votes meritocratically at size $k+l$.

Hence the majority is entrenched when it has size $k+l$. Note that this implies that $u_{k+l} = u_{k+l+1} = ... = u_{N-2} = 0$. This establishes the uniqueness of the super-entrenchment at level l within the class of equilibria such that a stronger majority makes more meritocratic recruitments. Furthermore, the argument implies that in any symmetric MPE in weakly undominated strategies, the majority is entrenched when it has size $k+l$.

J Endogenous candidacies: Proofs

A simple result used repeatedly in this section is the following: let $X(t)$ follow $\frac{dX}{dt} = \chi(-X + X^*)$ (with X^* the steady state value). Then $X(t) = (X(0) - X^*)e^{-\chi t} + X^*$, and the PDV of the flow $X(t)dt$

M and $M+1$. This gives that the weight of u_M on the LHS and the weight of u_{M-1} on the RHS are such that $u_{M-1} \leq u_M$ (as all other u_i are negative, and the sum of the flow payoffs is positive).

(weighted by time preference and exit probability) is a convex combination of the initial value and the steady state value: $\int_0^\infty e^{-(r+\chi)t}X(t)dt = \frac{1}{r+\chi}\left(\frac{(r+\chi)X(0)+\chi X^*}{r+2\chi}\right)$.

J.1 Endogenous candidacies when 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.

(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 14. ***(Only talented minority candidates have an outside option)*** *Assume (5) 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 (6), 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 (6), the majority selects $\sigma_1 = \sigma_0 = 1$.*

If (5) 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 5. *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,*

$$\frac{\partial V}{\partial M(T)}(M(T), S(T)) = b, \qquad \text{and} \qquad \frac{\partial V}{\partial S(T)}(M(T), S(T)) = s \tag{38}$$

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

$$\begin{aligned} M(t) &= \big[M(T) - x - (1-2x)\sigma_0^*\big]e^{-\chi(t-T)} + x + (1-2x)\sigma_0^*, \\ S(t) &= \big[S(T) - 2x\big]e^{-\chi(t-T)} + 2x \end{aligned}$$

Consequently,

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

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 (39) and (40), which are respectively the final time constraint

$$sS(T) - bM(T) = u - b \tag{39}$$

and the state dynamics

$$\frac{dM}{dt} = \chi\big[-M + x(\sigma_1 + 1 - \sigma_2) + (1-2x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS}{dt} = \chi\big[-S + x(\sigma_1 + \sigma_2)\big] \tag{40}$$

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 (39),

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

which together with (38) 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

$$\begin{cases} \dfrac{\partial H}{\partial \sigma_0} = \chi\left(e^{-(r+\chi)t} - \psi e^{-\chi(T-t)}\right)(1-2x)b, \\ \dfrac{\partial H}{\partial \sigma_1} = \chi e^{-(r+\chi)t}x(s+b) + \psi\chi e^{-\chi(T-t)}x(s-b), \\ \dfrac{\partial H}{\partial \sigma_2} = \chi e^{-(r+\chi)t}x(s-b) + \psi\chi e^{-\chi(T-t)}x(s+b) \end{cases} \tag{41}$$

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:

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

i.e. by using the final time constraint (39), 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 $\sigma_0 = 1$.[121]

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 (6) in the text – draws the line between regions 3 and 4.

(i) Suppose region 2 is reached at time $T < \infty$. Then (38) holds. The majority's optimization problem writes as

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

[121] 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.

subject to (42) and (43) which are respectively the final time constraint

$$sS(T) - bM(T) = u - b \tag{42}$$

and the state dynamics

$$\frac{dM}{dt} = \chi\big[-M + x\sigma_1 + (1-x)\sigma_0\big], \qquad \text{and} \qquad \frac{dS}{dt} = \chi\big[-S + x\sigma_1\big] \tag{43}$$

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 (42),

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

which together with (38) imply that

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

the Hamiltonian's partial derivatives write as

$$\begin{cases} \dfrac{\partial H}{\partial \sigma_0} = \chi(1-x)\Big(be^{-(r+\chi)t} - \psi b e^{-\chi(T-t)}\Big), \\ \dfrac{\partial H}{\partial \sigma_1} = \chi x\Big(e^{-(r+\chi)t}(s+b) + \psi\chi e^{-\chi(T-t)}(s-b)\Big) \end{cases} \tag{44}$$

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:

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

i.e. by using the final time constraint (42), 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 (4) 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 (5) 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 (6) in the text):

$$\int_0^T e^{-(r+\chi)t}(1-x)b\big[1-e^{-\chi t}\big]dt + \int_T^\infty e^{-(r+\chi)t}(1-x)b\big[1-e^{-\chi T}\big]e^{-\chi(t-T)}dt \\ \leq \int_T^{+\infty} e^{-(r+\chi)t}\big(3xs-u\big)\big[1-e^{-\chi(t-T)}\big]dt \tag{45}$$

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

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

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

$$\forall t \geq 0, \qquad u_t^{(4)} = [M_0 b + S_0 s - xs - b]e^{-\chi t} + xs + b,$$

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

$$\forall t \in [0, T], \qquad u_t^{(3)} = [M_0 b + S_0 s - xs - xb]e^{-\chi t} + xs + xb,$$

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

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

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

J.2 Endogenous candidacies: General exposition

We assume in the following that condition (5) is satisfied.

(A.2) Now 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 $\big(1/2, \dfrac{u - b/2}{s}\big)$ 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 [122]. 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[123], 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 select $\sigma_1 = \sigma_0 = 1$. 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 σ_0 such that region 2 is reached in a finite time, the lowest one is optimal.

As a consequence, 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 $\big(1/2, \frac{u - b/2}{s}\big)$, until this trajectory reaches the line defined by (6), after which the frontier is given as before by the latter, which is an increasing line parallel to region 2 [124]: 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 reaches the vertical axis before region 2, it subsequently goes up the vertical axis towards the point $(1/2, x)$. Since this point 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 Appendix J.3 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[125]. Namely, the organization engages in full affirmative action ($\sigma_0 = 0$) whenever it can, 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 $\big(1/2, \frac{u - b/2}{s}\big)$.

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

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

[124] 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 (6) then ensures that full affirmative action is optimal with respect to standard favoritism.

[125] 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. $(\big(1/2, \frac{u - b/2}{s}\big)$), 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[126]

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

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[127]

$$\int_0^{\overline{T}} e^{-(r+\chi)t}(1-x)b\big[1-\overline{\sigma}_0\big](1-e^{-\chi t})dt + \int_{\overline{T}}^{\infty} e^{-(r+\chi)t}(1-x)b\big[1-\overline{\sigma}_0\big]\big(1-e^{-\chi\overline{T}}\big)e^{-\chi(t-\overline{T})}dt \\ \leq \int_{\overline{T}}^{\infty} e^{-(r+\chi)t}(3xs-u)\big[1-e^{-\chi(t-\overline{T})}\big]dt \tag{46}$$

Given an initial state S_0 (and thus given $\overline{T}$), condition (46) 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 (46) with equality defines an upward-sloping curve in the plane (M, S), which we denote by Γ'. Moreover, since the LHS in (46) 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 $\big(1/2, \dfrac{u - b/2}{s}\big)$. 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 (6).[128]

(B) We now assume that both the majority's and the minority's talented candidates have the same (normalized) opportunity cost of 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 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

[126] 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

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

[127] See Appendix J.3 for details.

[128] 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 (6) then ensures that full affirmative action is optimal with respect to standard favoritism.

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. [129]

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 the next section (Appendix J.3) 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).

J.3 Proof of Proposition 7

Case A.2.b. 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 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 $(M,S) = \big(1/2, \frac{u-b/2}{s}\big)$. 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

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

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 $\big(1/2, \frac{u-b/2}{s}\big)$.
- 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

[129] 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.

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

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

The majority's optimization problem writes as

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

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[130] the optimal control σ_0

[130] 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:

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

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 $\big(\frac{1}{2}, \frac{u - b/2}{s}\big)$ 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 favoritism writes as

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

which yields (46) after rearranging.

Talented majority candidates have a participation constraint. Consider an "affirmative action" trajectory that reaches the interior of region 5 before the vertical axis. Then 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 (6) or (46) hold, and belongs to regions 4 or 5 otherwise.

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

which implies that:

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

$b/2$ upon joining the organization.

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

(B.2.a) If $u - b/2 \geq xs$, then region 5 and region 3 have no shared boundary[131]. Region 5 is given by the set of states below its boundary with region 4 (see case B.2.c below). Anticipating on B.2.c, region 5 is non-empty if and only if the initial state $(1/2, 0)$ satisfies (see (48) below)

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

i.e. if and only if

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

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 (46). Hence the boundary between region 5 and region 3 is given by the set of initial states (M_0, S_0) (satisfiying (46) with equality) such that

$$\begin{aligned}
&\int_0^{\overline{T}} e^{-(r+\chi)t}\left[[S_0 - x]\tilde{s}e^{-\chi t} + x\tilde{s} + [M_0 - x - (1-x)\overline{\sigma}_0]\tilde{b}e^{-\chi t} + x\tilde{b} + (1-x)\overline{\sigma}_0\tilde{b}\right]dt \\
&\quad + \int_{\overline{T}}^\infty e^{-(r+\chi)t}\left[\left(\tilde{u} - \frac{\tilde{b}}{2} - 2x\tilde{s}\right)e^{-\chi(t-\overline{T})} + 2x\tilde{s} + \left(\frac{\tilde{b}}{2} - (x + (1-2x)\sigma_0^*)\tilde{b}\right)e^{-\chi(t-\overline{T})} + [x + (1-2x)\sigma_0^*]\tilde{b}\right]dt \\
&= \int_0^\infty e^{-(r+\chi)t}\tilde{u}dt \qquad (47)
\end{aligned}$$

where $\overline{T} > 0$, $\overline{\sigma}_0 \in [0,1]$ are given whenever they exist[132] by

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

The LHS in (47) 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

[131] 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 $\big(1/2, \frac{u - b/2}{s}\big)$.]

[132] 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

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

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.[133]. Therefore, the frontier between regions 3 and 5 has a decreasing slope in the plane (M, S) 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 (46)[134], then region 5 is non-empty if it includes the state $(1/2, 0)$, i.e. if

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

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 $xs > 0$):

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

[133]Indeed, explicit computations yield

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

Similarly,

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

i.e. after rearranging,

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

Therefore, since $u - b/2 < xs$, $u < 3xs$, $S_0 < x$, and $M_0 < e^{\chi\overline{T}}\left[1/2 - x\right] + x$, 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 $\overline{T} \gg 1$.

[134]$(1/2, 0)$ satisfies (46) if and only if

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

i.e. if and only if

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

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

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

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, 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 (46), then a necessary and sufficient condition for region 5 to be non-empty is given by the condition stated in 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.

(B.2.c) The frontier between regions 5 and 4 is given by the set of states (violating (46) if $u - b/2 < xs$) such that

$$\int_0^\infty e^{-(r+\chi)t}\Big[[S_0 - x]\tilde{s}e^{-\chi t} + x\tilde{s} + [M_0 - 1]\tilde{b}e^{-\chi t} + \tilde{b}\Big]dt = \int_0^\infty e^{-(r+\chi)t}\tilde{u}dt \tag{48}$$

Since the LHS in (48) 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

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

i.e. if

$$u > b + \frac{\chi}{r + 2\chi}xs \tag{49}$$

Hence, if (49) 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 (49) 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 $(1, x)$.[135]

K Proof of Proposition 8

The dynamics of quality dominance. If (7) is satisfied at date 0, then $[S_1 - S_2]$ converges towards $2x$, while $[M_2 - (1 - M_1)]$ converges towards $(1 - x)$. (Recall that (7) is the non-profitability condition for a collective deviation by all talented B-candidates from joining organization 1 to joining organization 2). Hence if (7) is satisfied at time 0, then by convexity, it is satisfied at any later time $t > 0$ if and only if

[135] 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 ($xs + b > u$).

the steady state satisfies (7), i.e. if and only if[136]

$$2xs - (1-x)b \geq \frac{\chi}{r+\chi}\big[(1-x)b - xs\big], \tag{50}$$

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

If (50) 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).[137]

By contrast, if (7) and (50) hold, then whenever the initial state verifies (7), $[S_1 - S_2]$ converges towards $2x$, while $[M_2 - M_1]$ converges towards x: there is quality dominance in the long run. In line with the rest of this section, *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 "comparative advantage" of organization 1 with respect to organization 2 from the perspective of (talented) B-group candidates. Condition (7) can thus be written as

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

If (7) and (50) 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 (50) 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

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

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

[137] Indeed, if (50) is violated, then there exists a later (finite) time at which (7) 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.

co-existence of two increasing-dominance equilibria[138]

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

i.e. if and only if

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

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,

- 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[139]

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

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.

[138] 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).

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

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

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[140], 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 link for such a deviation)[141]

$$\big[S_2(0) - S_1(0)\big]s + \big[M_1(0) + M_2(0) - 1\big]b \geq -\frac{\chi}{r+\chi}(1-2x)b \tag{51}$$

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 (51) 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 (51) converges to $2xs + (1-x)b > 0$ [142], 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.

L Proof of Proposition 9

We first show that meritocratic equilibrium *strategies* are no longer so when candidates reapply, for s/b in some interval $[1, \rho^{\mathrm{m}} + \epsilon)$ with $\epsilon > 0$. We then show that the meritocratic equilibrium *path* starting from an initial state with empty storage is no longer an equilibrium path for s/b in some interval $[1, \rho^{\mathrm{m}} + \epsilon)$ with $\epsilon > 0$: an equilibrium may be observationally equivalent to a meritocratic equilibrium by exhibiting the same recruitment path, without necessarily be meritocratic off the equilibrium path (more on this below).

We define the meritocratic equilibrium as an equilibrium in which the majority always recruits the best candidate available[143] for any stocks of candidates, and look for necessary conditions for the meritocratic equilibrium to exist. We show the latter are more often binding when candidates reapply than when they cannot. Namely, when candidates reapply, we exhibit one deviation that is profitable for s/b

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

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

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

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

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.

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

[143] Namely the best candidate among current-period and stored candidates, breaking ties in favour of in-group candidates as before.

a bit above ρ^{m} (and for all $s/b \in [1, \rho^{\mathrm{m}}]$). Note that we do not derive a sufficient condition for existence.

Two effects (which we will successively illustrate) are at play, shrinking the existence region of meritocracy: (i) the ability to recall a talented minority candidate increases the value of entrenchment; and (ii) the preferential treatment given by the majority to its in-group talented candidate(s) in store makes an incumbent majority with a large number of talented minority candidates in store less willing to relinquish control.

To illustrate both forces at play, consider first $x = 1/2$ (so that $\rho^{\mathrm{m}} = 1$), and $s/b = 1$. Suppose the majority has size k, and no talented majority candidate available[144] but an infinite number of talented minority ones in store. Recruiting a talented minority candidate instead of an untalented majority one gives a differential payoff equal to

$$s - b + \delta \frac{k-1}{N-1}\left(\frac{s}{1-\delta} - V_{k+1,0,\infty}\right) = \delta \frac{k-1}{N-1}\left(\frac{s}{1-\delta} - V_{k+1,0,\infty}\right)$$

where $V_{k+1,0,\infty}$ is the majority value function when it has size $k+1$, no talented majority candidate in store and an infinite number of talented minority ones in store. Since for $x = 1/2$, a majority with size $k+1$ can secure in each period an (expected) flow quality payoff equal to $\tilde{s}$, and for at least the first two periods, an (expected) flow homophily payoff equal to $\tilde{b}/2$ [145], we have that $V_{k+1,0,\infty} > s/(1-\delta)$. Furthermore, as the majority cannot do better than $\tilde{s}$ in terms of flow quality payoff, the term $[s/(1-\delta) - V_{k+1,0,\infty}]$ does not decrease with s, but strictly decreases with b. Therefore, the above differential payoff is strictly negative for any s/b in an upper neighbourhood of 1. Because of time discounting ($\delta_0 < 1$), the result holds when the majority has in store a sufficiently large finite number of talented minority candidates. Hence, for $x = 1/2$, there exists a strictly profitable deviation away from meritocracy for $s/b \in [\rho^{\mathrm{m}}, \rho^{\mathrm{m}} + \epsilon)$.

Consider now $x < 1/2$ (so that $\rho^{\mathrm{m}} > 1$), and $s/b = \rho^{\mathrm{m}}$. A necessary condition for the meritocratic equilibrium to exist is that a repeated deviation towards entrenchment whenever the majority is tight ($M = k$) and has no talented majority candidate available and exactly one talented minority candidate available, be non profitable. Upon permanently deviating to entrenchment, the majority has one talented minority candidate in store, and either size k or $k+1$. Yet, for $x < 1/2$, an *entrenched* majority's value function strictly increases with the number of talented minority candidates in store[146]. Hence, when candidates reapply, a permanent deviation away from meritocracy becomes more profitable. Furthermore, an inspection of the additional payoff due to storability shows that the latter increases with s and decreases with b. Intuitively, this derives from the fact that having a talented minority candidate in store leads to the latter being recruited (at some point, with strictly positive probability) instead of a (talented or untalented) in-group candidate or an untalented out-group candidate, thus yielding a positive quality gain and a positive homophily loss with respect to the payoff when candidates cannot

[144] Namely, it has no such candidate in store, and the current-period majority candidate is untalented.

[145] In particular, reverting to the meritocratic strategy yields to the current majority group an (expected) flow payoff equal to $\tilde{s} + \tilde{b}/2$ as long as it retains control over the organization, and $\tilde{s}$ after it has relinquished it to the other group.

[146] Indeed, an entrenched majority solves an optimal control problem. Moreover, as $x < 1/2$, the majority faces two untalented current-period candidates with a strictly positive probability ($1 - 2x > 0$), in which case, whenever it is not tight ($M > k$) and whenever it has a talented minority candidate in store, it recruits the latter, thus receiving a strictly positive differential payoff with respect to the empty-storage state. Indeed, the differential payoff from recruiting a stored talented minority candidate instead of an untalented majority candidate whenever the majority is not tight, is bounded below by:

$$s - b - x(s-b)\frac{\delta k/(N-1)}{1 - \delta k/(N-1)} > (1-x)(s-b) > 0$$

reapply. Therefore, since in the absence of storability, we have the equivalence between the profitability of one-shot and permanent deviations[147], there exists a profitable deviation away from meritocracy for $s/b > \rho^{\mathrm{m}}$ (and for all $s/b \in [1, \rho^{\mathrm{m}}]$) , i.e. the existence region of meritocracy shrinks.

Finally we show that the meritocratic equilibrium *path* starting from an initial state with empty storage is no longer an equilibrium path for s/b in some interval $[\rho^{\mathrm{m}}, \rho^{\mathrm{m}} + \epsilon)$ with $\epsilon > 0$. We first note that, on the meritocratic equilibrium path starting from an initial state with empty storage, storage is never used[148]. Hence, considering the repeated deviation to entrenchment described above yields that, for $x < 1/2$, there exists a strictly profitable deviation away from this equilibrium path for s/b slightly above ρ^{m} (and for all $s/b \in [1, \rho^{\mathrm{m}}]$). As a consequence, when $x < 1/2$, then for s/b in some interval $[\rho^{\mathrm{m}}, \rho^{\mathrm{m}} + \epsilon)$ with $\epsilon > 0$, the meritocratic equilibrium path starting from an initial state with empty storage is no longer so.

M Proof of Proposition 11

We first show the validity of the remark in the text on a blind principal ($\lambda = 0$), before establishing Proposition M.

If the principal does not observe horizontal types and in particular the majority size, it worsens the efficiency of its interventions as it cannot fine-tune its interventions. Hence if it is an equilibrium for the principal not to intervene when it observes horizontal types, it is also an equilibrium to do so when the principal is totally blind. We thus consider the case where the principal observes horizontal types and show that, taking as given members' beliefs on the principal's strategy, it is optimal for the latter not to engage in interventions. There is clearly no benefit for the principal to intervene whenever the majority is not tight ($M \geq k + 1$) – or whenever it is tight and meritocratic – as then the majority's choice maximizes the organization's quality and, by resolving ties in favour of the majority candidate, also maximizes the homophily payoff conditional on maximizing the organization's quality. Hence, for $s > b$ and $q \geq 1$, the majority's choice is optimal from the principal's point of view.[149]

Thus we now need to show that it is optimal[150] for the principal not to intervene in the entrenchment equilibrium when majority is tight ($M = k$). Since a tight entrenched majority always votes for its own candidate, its vote carries no information on the candidates' respective talents: the principal cannot do better by observing horizontal types than it can without. Hence, from the quality perspective, the principal picks the (of "a" if there is a tie) right candidate with probability 1/2, whereas the majority does so with probability $(1 - x) \geq 1/2$. Similarly, the majority takes the homophily-maximizing decision with probability 1, while the principal can at best replicate this probability if it observes horizontal types, and can only do so with probability 1/2 if it does not. Hence the principal cannot outperform the majority's decision.

[147] Hence, when candidates cannot reapply, the above repeated deviation yields a zero differential payoff for $s/b = \rho^{\mathrm{m}}$.

[148] Indeed, as we assume $\alpha = 0$, the organization faces at most one new talented candidate each period, and on the meritocratic equilibrium path, recruits her/him.

[149] Fix $s > b$. Since the quality payoff accrues to all members of the organization, while the homophily benefit only accrues to the in-group members, this optimality persists for q in a lower neighbourhood of 1. Furthermore, the neighbourhood expands toward 0 as the ratio s/b increases.

[150] Strictly so if there is any small cost of intervention, or if the principal internalizes members' homophily benefits.

We now turn to the proof of Proposition 11.

Proof of claim (i). Let $\lambda > 0$ be the probability that the principal learns the quality of the candidates. The proof unfolds in two steps:

(a) We show that for s/b sufficiently close to 1, there exists a profitable deviation from canonical entrenchment in $k+1$ (the unique outcome when s/b is close to 1 and $\lambda = 0$) toward super-entrenchment at level 1. The argument then extends to full-entrenchment.

(b) We show that for s/b sufficiently close to 1, there can be no profitable deviation from full entrenchment.

(a). For $i \geq k$, let V_i be the majority value function in the canonical entrenchment equilibrium with probability of intervention $\eta = x\lambda > 0$. In order to alleviate the notation, we drop the superscript e and the notation for the dependence on λ. Consider a deviation from canonical entrenchment to super-entrenchment in $k+1$, i.e. the majority voting its own, less talented candidate against the strictly more talented minority one, and being overruled with probability λ. The (one-shot) differential payoff from the deviation at $M = k+1$ writes

$$\begin{aligned}\Delta &\equiv (1-\lambda)\left[b - s + \delta\left(\frac{k+1}{N-1}V_{k+1} + \frac{k-2}{N-1}V_{k+2}\right) - \delta\left(\frac{k}{N-1}V_k + \frac{k-1}{N-1}V_{k+1}\right)\right] \\ &= (1-\lambda)\left[b - s + \delta\left(\frac{k-2}{N-1}u_{k+1} + \frac{k}{N-1}u_k\right)\right]\end{aligned}$$

where $u_i = V_{i+1} - V_i$. The sequence $(u_i)_{1\leq i\leq N-2}$ satisfies Equation (10) for any $i \geq k+1$, and Equation (12) for any $i \leq k-3$, while

$$\begin{cases} \left[1 - \delta(1-x)\frac{k}{N-1} - \delta x\lambda\frac{k-1}{N-1}\right]u_k = x(1-\lambda)(s-b) + \delta(1-x)\frac{k-2}{N-1}u_{k+1} + \delta x\lambda\frac{k-1}{N-1}u_{k-1} \\ \left[1 - \delta(1-x\lambda)\right]u_{k-1} = (1-2x\lambda)b + \delta(1-x\lambda)\left[\frac{k-2}{N-1}u_{k-2} + \frac{k-1}{N-1}u_k\right] \\ \left[1 - \delta(1-x)\frac{k+1}{N-1} - \delta x\lambda\frac{k-2}{N-1}\right]u_{k-2} = -x(1-\lambda)(s+b) + \delta(1-x)\frac{k-3}{N-1}u_{k-3} + \delta x\lambda\frac{k}{N-1}u_{k-1} \end{cases} \tag{52}$$

Summing up on all indices yields[151]

$$\left[1 - \delta\frac{x}{N-1} - \delta(1-x)\right](u_1 + u_{N-2}) + \big(1-\delta\big)\sum_{i=2}^{N-3} u_i = (1-2x)b > 0 \tag{53}$$

Fix $b > 0$. For any $s \geq b$, the same argument as the one used in the proof of Lemma 1 yields $u_k > u_{k+1} > ... > u_{N-2} > 0$. Put succinctly, one supposes by contradiction that $u_{N-2} \leq 0$ and reaches a contradiction showing by induction, using (10) together with the above system, that this implies $u_{k-1} \leq 0$. Then, if $u_1 \leq 0$, (12) implies $u_i \leq 0$ for all i, which contradicts (53); whereas if $u_1 > 0$, (12) implies $u_{k-1} > 0$ and we reach again a contradiction. Hence $u_{N-2} > 0$ and the same induction argument using (10) thus brings the result.

[151] Assuming $k \geq 4$. The expression for $k \in \{2,3\}$ writes differently on the LHS but has the same implication.

The differential deviation payoff is thus strictly positive if and only if

$$\delta\left(\frac{k-2}{N-1}u_{k+1}+\frac{k}{N-1}u_k\right)>s-b \tag{54}$$

Consequently, for $s=b$, (54) is satisfied as it writes

$$\delta\left(\frac{k-2}{N-1}u_{k+1}+\frac{k}{N-1}u_k\right)>0$$

Lastly, since for fixed b, $(u_i)_i$ is continuous with respect to s, this implies that for any s/b sufficiently close to 1, there exists a strictly profitable (one-shot) deviation from canonical entrenchment to super-entrenchment.

As a by-product of the proof, we have by the same argument that whenever $\eta=0$, there exists no profitable deviation from canonical entrenchment to super-entrenchment as then[152]

$$\delta\left[\frac{k-2}{N-1}u_{k+1}(\lambda=0)+\frac{k}{N-1}u_k(\lambda=0)\right]<\frac{x}{1-x}\left[\left(1-\delta(1-x)\frac{k}{N-1}\right)^{-1}-1\right](s-b)<s-b$$

The same argument shows that, for s/b sufficiently close to 1, there exist profitable deviations from any level $l\geq 0$ of entrenchment toward entrenchment at a higher level, and thus in particular toward full-entrenchment.

Lastly, we argue that this establishes the uniqueness of the full-entrenchment equilibrium among all symmetric MPEs in weakly undominated strategies. To this end, we show that, for s/b in a neighbourhood of 1, any symmetric MPE in weakly undominated strategies is monotonic, in the sense that a stronger majority makes more meritocratic recruitments. The result crucially relies on the fact that the minority is never pivotal – as opposed to the absenteeism (Section 3.2) and supermajority rules (Section 6.3) settings.

Let $s=b>0$. We show that in any symmetric MPE in weakly undominated strategies, the sequence of differential value function $(u_M)_{M\geq k-1}$ is strictly positive and strictly decreases with M. As a consequence, by continuity, it is so for s/b in a neighbourhood of 1, and this in turn implies that, for s/b in such a neighbourhood, any symmetric MPE in weakly undominated strategies is monotonic as deviation differential payoffs at majority size M towards meritocracy write as

$$s-b-\delta\left[\frac{M-1}{N-1}u_{M-1}+\left(1-\frac{M}{N-1}\right)u_M\right]$$

and are thus increasing with M.

[152] Indeed,

$$\left[1-\delta(1-x)\frac{k}{N-1}\right]u_k(\lambda=0)=x(s-b)+\delta(1-x)\left(1-\frac{k+1}{N-1}\right)u_{k+1}(\lambda=0)<x(s-b)$$

For $s = b > 0$, we have that

$$\begin{cases} u_{k-1} = (1-2x\lambda)b + \delta(1-x\lambda)\left[\frac{k-2}{N-1}u_{k-2} + u_{k-1} + \frac{k-1}{N-1}u_k\right] & \text{in an equilibrium in which the majority is entrenched in } k, \\ u_{k-1} = (1-2x)b + \delta(1-x)\left[\frac{k-2}{N-1}u_{k-2} + u_{k-1} + \frac{k-1}{N-1}u_k\right] & \text{in an equilibrium in which it is meritocratic in } k \end{cases}$$

and for any majority size $M \leq N-2$,

$$\begin{cases} u_M = \delta(1-x\lambda)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\lambda\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ \qquad \text{in an equilibrium in which the majority is entrenched in } M, M+1, \\ u_M = \delta(1-x)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ \qquad \text{in an equilibrium in which the majority is meritocratic in } M, M+1, \\ u_M = \delta(1-x)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\lambda\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ \qquad \text{in an equilibrium in which the majority is entrenched (resp. meritocratic) in } M(\text{resp. } M+1), \\ u_M = \delta(1-x\lambda)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\left[\frac{M-1}{N-1}u_{M-1} + u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] \\ \qquad \text{in an equilibrium in which the majority is meritocratic (resp. entrenched) in } M(\text{resp. } M+1), \end{cases}$$

together with similar expressions for u_i when $i \leq k-2$.

Hence, we apply the usual argument: supposing by contradiction that $u_{N-2} \leq 0$, and working by induction using the sums of u_i over appropriate indices in order to reach the contradiction – which ultimately derives from the fact that there is a unique flow differential payoff, and that it is equal either to $(1-2x\lambda)b$ or $(1-2x)b$, which are both strictly positive. This yields that in any equilibrium, $u_{N-2} > 0$, and the above system then implies that $u_i > 0$ for all $i \in \{k-1, ..., N-2\}$ as was to be shown.

(b). We now show existence, i.e. that for s/b sufficiently close to 1 there can be no profitable deviation from full entrenchment. The argument is analogous to the one just used. In order to alleviate the notation, we again omit the superscript and the dependence on η and simply write V for the value function and u for its first difference.

The deviation differential payoff from full-entrenchment to entrenchment at a lower level in $M = N-1$ whenever the minority candidate is more talented writes

$$\Delta \equiv (1-\lambda)\left[s - b - \delta\frac{N-2}{N-1}u_{N-2}\right]$$

Explicit computation with (8)-(9) yield:

$$u_{N-2} = \delta(1-x\lambda)\frac{N-2}{N-1}u_{N-2} + \delta x\lambda\left[\frac{N-3}{N-1}u_{N-3} + \frac{1}{N-1}u_{N-2}\right]$$

and more generally for any $M \geq k$,

$$u_M = \delta(1-x\lambda)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\lambda\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right]$$

while for any $i \leq k-2$,

$$u_i = \delta(1-x\lambda)\left[\frac{i-1}{N-1}u_{i-1} + \left(1-\frac{i}{N-1}\right)u_i\right] + \delta x\lambda\left[\frac{i-1}{N-1}u_i + \left(1-\frac{i+1}{N-1}\right)u_{i+1}\right]$$

with

$$\left[1-\delta\big(1-x\lambda\big)\right]u_{k-1} = (1-2x\lambda)b + \delta(1-x\lambda)\left[\frac{k-1}{N-1}u_k + \frac{k-2}{N-1}u_{k-2}\right]$$

Summing up over all indices yields

$$\left[1-\delta\left(1-x\lambda\frac{N-2}{N-1}\right)\right]u_{N-2} + \left[1-\delta\left(1-\frac{x\lambda}{N-1}\right)\right]u_1 + (1-\delta)\sum_{i=2}^{N-3}u_i = (1-2x\lambda)b > 0 \qquad (55)$$

Fix $b > 0$ and let $s = b$. The usual argument implies that $u_{N-2} > 0$. Indeed, if not, then the above equations imply by induction that $u_k \leq u_{k+1} \leq ... \leq u_{N-2} \leq 0$ and thus $0 \geq u_1 \geq u_2 \geq ... \geq u_{k-1}$, which yields to a contradiction with (55). Therefore, $u_{N-2} > 0$, and by induction again $u_k > u_{k+1} > ... > u_{N-2} > 0$. Hence the differential deviation payoff when the majority has size $N-2$ writes for $s = b$ as

$$\Delta = -(1-\lambda)\delta\frac{N-2}{N-1}u_{N-2} < 0$$

The result for $s = b$ obtains by noting that since $u_k > u_{k+1} > ... > u_{N-2} > 0$, the one-shot deviation when majority has size $N-1$ is the most profitable one-shot deviation from the full-entrenchment strategy. The result then extends to s/b in a neighbourhood of 1 by continuity.

Proof of claim (ii). The principal cannot expand the existence region of meritocracy by its interventions as the prospect of its overruling a majority's decision only scales down (by a strictly positive factor) the one-shot deviation differential payoff from meritocracy to entrenchment. Hence, under our assumption that the meritocratic equilibrium is selected whenever it exists, the principal fails to expand the region where one should expect meritocracy.

Whenever informed, the principal has a profitable (one-shot) deviation from no-intervention. Hence, if the principal cannot commit, majority members anticipate the principal steps in whenever informed. Proposition 11 then implies that an entrenched organization at best remains (canonically) entrenched or meritocratic, and otherwise goes super-entrenched – and most notably fully-entrenched for s/b in a non-empty neighbourhood of 1.

Hence in particular, for s/b sufficiently close to 1, the organization is fully-entrenched. Since the principal is only informed with probability strictly below 1, it cannot compensate all the "un-meritocratic" decisions made by the organizations. Hence, at any majority size $M \geq k+1$, the principal would be better off in terms of flow-payoffs, if it could commit not to intervene.

By contrast, whenever the majority is tight, entrenchment would have prevailed, and so the principal may find it optimal to intervene. Fix a probability $\lambda \in (0,1)$ of the principal's being informed, and consider any s/b sufficiently close to 1 such that, given λ, the unique equilibrium is full entrenchment. Suppose the principal values only ergodic efficiency. Then it would be better off committing not to

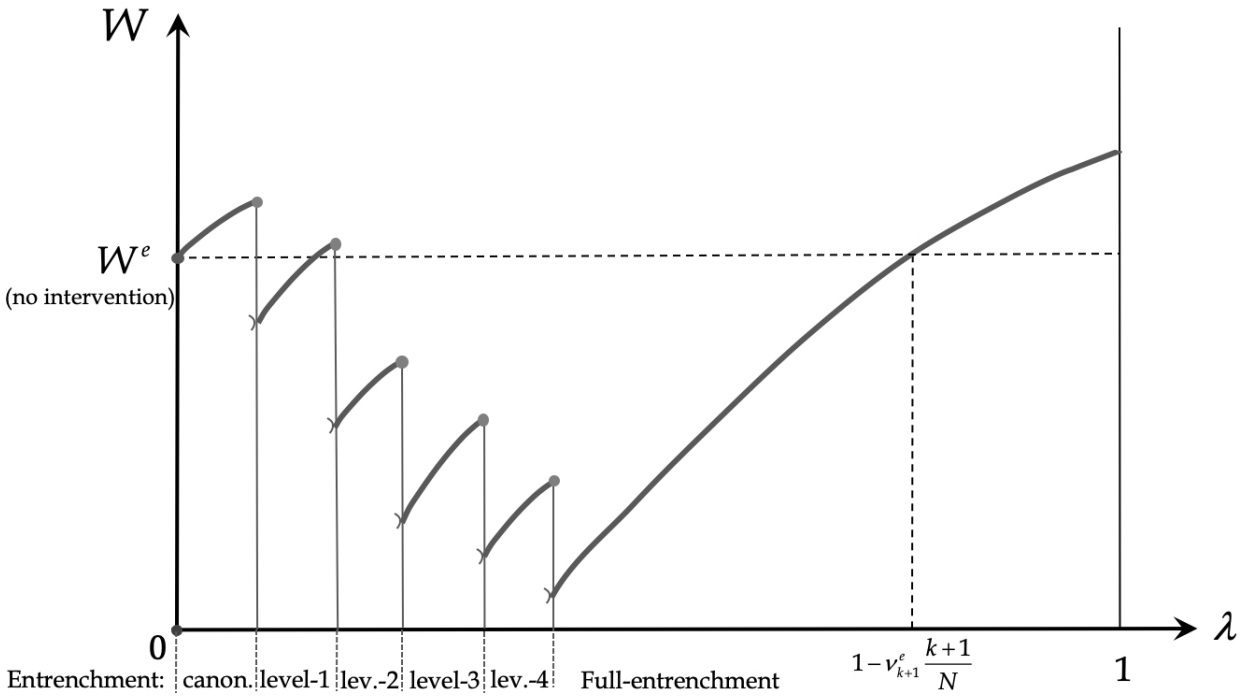

Figure 5: Principal's ergodic aggregate welfare as a function of λ (for $N = 12$).

intervene if and only if

$$N(N-1)(1-\lambda)xs > N(N-1)\nu^{\mathrm{e}}_{k+1}\frac{k+1}{N}xs, \qquad \text{i.e.} \qquad \lambda < 1-\nu^{\mathrm{e}}_{k+1}\frac{k+1}{N}$$

Lastly, Figure 5 depicts the principal's ergodic aggregate payoff as a functionits probability λ of being informed for fixed $s/b < \rho^{\mathrm{e}}$, assuming equilibrium selection and that the principal values only ergodic aggregate quality.

N Proof of Proposition 12

Consider an entrenched organization. Because of the Pareto-dominance selection (meritocracy prevails whenever it exists as an equilibrium), $s/b < \rho^{\mathrm{m}}$. Let $T \equiv \eta y$ denote equal the minimal expected bonus per member needed for the organization to move from entrenchment to meritocracy[153] For the sake of exposition, assume the principal does not value members' homophily benefits, and thus letting ξ be the cost of public funds[154], the principal's objective function writes as the ergodic welfare with per-period welfare given by $W = qS - \xi T$ [155]. Note that such an objective constitutes an *upper* bound on the admissible cost of a policy as homophily decreases when the organization goes from entrenchment to

[153]Namely,

$$\frac{s^+(\eta,y)}{b} = \rho^{\mathrm{m}}, \qquad \text{i.e.} \qquad \eta y = \left(\frac{b}{s}\rho^{\mathrm{m}} - 1\right)\tilde{s} > 0$$

[154]The interpretation of ξ depends on the principal's welfare objective. If it is solely concerned with maximizing the (ergodic aggregate) quality of the organization, then ξ is the total cost of intervention, i.e. the sum of the payment and its shadow cost. By contrast, if the principal internalizes the "material" welfare of members, i.e. the sum of their quality payoffs and (possibly) rewards for quality (as opposed to their non-material welfare which consists of homophily benefits), then ξ is only the shadow cost of public funds.

[155]This objective may be interpreted as the limit of the main objective for $q, \xi \to \infty$.

meritocracy (see Section 2.2.2). From previous computations on ergodic welfare, the (ergodic) efficiency gain from disentrenchment writes as $S^{\mathrm{m}} - S^{\mathrm{e}} = N(N-1)\nu^{\mathrm{e}}_{k+1}\frac{k+1}{N}x\frac{\tilde{s}}{1-\delta} > 0$. Rewarding quality is thus optimal for the principal if and only if

$$\xi\eta y N^2(\overline{x}+x) \leq N(N-1)\nu^{\mathrm{e}}_{k+1}\frac{k+1}{N}x\tilde{s}$$

where $N[\overline{x}+x]$ is the average number of talented members in a meritocratic organization, and ν^{e}_{k+1} the objective ergodic probability of majority size $k+1$ in the entrenched equilibrium (see Section 2.2.2). The above inequality rewrites as a condition on the administrative cost of public funds:[156]

$$\xi \leq \frac{(k+1)(N-1)}{N^2}\cdot\frac{x\nu^{\mathrm{e}}_{k+1}}{\overline{x}+x}\cdot\frac{\tilde{s}}{\eta y} = \frac{(k+1)(N-1)}{N^2}\cdot\frac{x\nu^{\mathrm{e}}_{k+1}}{\overline{x}+x}\cdot\frac{\frac{s}{b}}{\rho^{\mathrm{m}}-\frac{s}{b}}$$

Note that the RHS strictly increases with s/b and goes to $+\infty$ as s/b goes to ρ^{m} [157]. The result follows.

The same argument applies if the principal's objective writes as $W = qS + B - \xi T$, yielding a higher threshold ρ_ξ (as $B^{\mathrm{m}} < B^{\mathrm{e}}$).

O Proof of Proposition 13

O.1 Proof of Proposition 13-(i): Affirmative action expands the existence region of meritocracy

The result is (almost) immediate[158] for a representation threshold of 1. For the sake of exposition and in order to get the spirit of the proof, we first focus on a representation threshold of $k-1$, whereby the minority size at the end of any period must be at least equal to $k-1$, before turning to the general proof for any representation threshold $R \in \{1, ..., k-1\}$.

Case $R = k-1$. We first show that the representation constraint strictly decreases the lower bound of the existence region of the meritocratic equilibrium, denoted by ρ^{m}_{AA}. We then prove that the existence region of meritocracy writes as $[\rho^{\mathrm{m}}_{AA}, +\infty) \supset [\rho^{\mathrm{m}}, +\infty)$. Because we focus on the existence of meritocratic equilibria, we henceforth omit the superscript "m".

For any $i \in \{1, ..., N-1\}$, let V_i (resp. for any $i \in \{k-2, ..., k+1\}$, $\tilde{V}_i$) denote the value function of being in a group of size i in the meritocratic equilibrium in the baseline model (resp. with a representation threshold of $k-1$). We first show that $\tilde{V}_{k-1} - \tilde{V}_{k+1} > V_{k-1} - V_{k+1}$, i.e. letting for any i, $u_i \equiv V_{i+1} - V_i$

[156] By Inequality (23), a lower bound on the RHS of the above equation is given by

$$\frac{(k+1)(N-1)}{N^2}\cdot\frac{x\nu^{\mathrm{e}}_{k+1}}{\overline{x}+x}\cdot\frac{\tilde{s}}{\eta y} \geq \frac{(k+1)(N-1)^2}{(k-1)N^2}\cdot\frac{x(1-2x)\nu^{\mathrm{e}}_{k+1}}{\overline{x}+x}\cdot\frac{(1-\delta)}{\delta}$$

[157] The monotonicity of the RHS with respect to N is non-trivial. Namely, although the first two terms decrease with $N \geq 4$, so that $(k+1)(N-1)\nu^{\mathrm{e}}_{k+1}/N^2$ decreases with N, the comparative statics of ρ^{m} with respect to N are non-trivial. Nonetheless, for N large, the first two terms $(k+1)(N-1)\nu^{\mathrm{e}}_{k+1}/N^2$ are in $O(1/N)$, while for $\delta_0 < 1$, ρ^{m} is in $0(1)$. Therefore, the RHS is in $0(1/N)$ for N large, which is intuitive: the upper bound on the admissible cost of public funds is inversely proportional to the size of the organization, i.e. to the number of individuals to whom the bonus must be distributed.

[158] The argument is significantly shorter in this case than with $R \geq 2$ since the minority's value function in the canonical entrenched equilibrium writes as in the baseline model with no affirmative action (due to the conditioning on still being a member next period).

and $\tilde{u}_i \equiv \tilde{V}_{i+1} - \tilde{V}_i$, that $\tilde{u}_k + \tilde{u}_{k-1} - (u_k + u_{k-1}) < 0$.

The representation threshold is binding when the majority has size $k+1$ at the beginning of a period, or equivalently when the minority has size $k-2$. So the majority must coopt the out-group candidate. Intuitively, this lowers the valuation for the majority when $M = k+1$ and increases it for the minority ($i = k-2$). Formally,

$$\begin{cases} \tilde{V}_{k+1} = \overline{x}s + \delta\left[\dfrac{k}{N-1}\tilde{V}_k + \dfrac{k-1}{N-1}\tilde{V}_{k+1}\right] \\ \tilde{V}_{k-2} = \overline{x}s + b + \delta\left[\dfrac{k-2}{N-1}\tilde{V}_{k-2} + \dfrac{k+1}{N-1}\tilde{V}_{k-1}\right] \end{cases}$$

Hence, by using (8)-(9) and the monotonicity of the value function (Lemma 1, whose proof also applies to affirmative action environment),[159]

$$\begin{cases} \left[1 - \delta\dfrac{k-1}{N-1}\right]\left(V_{k-1} - \tilde{V}_{k-1}\right) > xs + (1-x)b + \delta\dfrac{k}{N-1}\left(V_k - \tilde{V}_k\right) \\ \left[1 - \delta\dfrac{k-2}{N-1}\right]\left(\tilde{V}_{k-2} - V_{k-2}\right) > -xs + (1-x)b + \delta\dfrac{k+1}{N-1}\left(\tilde{V}_{k-1} - V_{k-1}\right) \end{cases}$$

Since both V_i and $\tilde{V}_i$ satisfy (8)-(9) for $i \in \{k-1, k\}$, computations yield that

$$\begin{cases} \left[1 - \delta x\dfrac{k-1}{N-1}\right]\left(\tilde{u}_k - u_k\right) < -xs - (1-x)b + \delta x\dfrac{k-1}{N-1}\left(\tilde{u}_{k-1} - u_{k-1}\right) \\ \left[1 - \delta x\dfrac{k-2}{N-1}\right]\left(\tilde{u}_{k-2} - u_{k-2}\right) < xs - (1-x)b + \delta x\dfrac{k}{N-1}\left(\tilde{u}_{k-1} - u_{k-1}\right) \end{cases} \tag{56}$$

Because (13) (which relates u_{k-1} to u_{k-2} and u_k) also applies to the affirmative action environment,

$$\left[1 - \delta(1-x)\right]\left(\tilde{u}_{k-1} - u_{k-1}\right) = \delta(1-x)\left[\frac{k-1}{N-1}\left(\tilde{u}_k - u_k\right) + \frac{k-2}{N-1}\left(\tilde{u}_{k-2} - u_{k-2}\right)\right],$$

(56) implies that: $\tilde{u}_{k-1} - u_{k-1} < 0$, and thus by (56) again, that: $\tilde{u}_k - u_k < 0$, which yields the result.

Consequently, for any s, b, a (one-shot) deviation from meritocracy to entrenchment when the majority has size k is strictly less profitable with a representation threshold of $k-1$ than without:

$$s - b + \delta\frac{k-1}{N-1}(\tilde{V}_{k-1} - \tilde{V}_{k+1}) > s - b + \delta\frac{k-1}{N-1}(V_{k-1} - V_{k+1})$$

The above computations together with the usual argument further imply that the LHS in the above inequality is linear in s and b, decreases[160] with b, increases with s, and thus increases with s/b. Therefore,

[159] Namely,

$$\begin{cases} \left[1 - \delta\dfrac{k-1}{N-1}\right]\left(V_{k-1} - \tilde{V}_{k-1}\right) = xs + (1-x)b + \delta\dfrac{k}{N-1}\left(V_k - \tilde{V}_k\right) + \delta(1-x)\left[\dfrac{k-2}{N-1}u_{k+1} + \dfrac{k}{N-1}u_k\right] \\ \left[1 - \delta\dfrac{k-2}{N-1}\right]\left(\tilde{V}_{k-2} - V_{k-2}\right) = -xs + (1-x)b + \delta\dfrac{k+1}{N-1}\left(\tilde{V}_{k-1} - V_{k-1}\right) + \delta(1-x)\left[\dfrac{k-3}{N-1}u_{k-3} + \dfrac{k+1}{N-1}u_{k-2}\right] \end{cases}$$

[160] Indeed, note that the omitted terms in (56) – i.e. the negative terms replaced by their upper bound of 0 –, write for the first equation as

$$-\delta(1-x)\left[\frac{k+l-1}{N-1}u_{k+l-1} + \frac{k-l-1}{N-1}u_{k+l}\right],$$

which is thus proportional to $(-b)$ (see proof of Lemma 1 for details). Similarly for the second equation.

denoting by ρ^{m}_{AA} the value of s/b for which the LHS is equal to 0, we have that $\rho^{\mathrm{m}}_{AA} < \rho^{\mathrm{m}}$.

Lastly, the same argument as in the proof of Proposition 1 (cf. Section C.2) yields that (a) for any $s/b \geq \rho^{\mathrm{m}}_{AA}$, (constrained) *meritocratic choices* are an equilibrium, with ties possibly broken in favour of the minority candidate, and that (b) for any $s/b < \rho^{\mathrm{m}}_{AA}$, there can be no meritocratic recruitment in equilibrium when the majority is tight. Consequently, the existence region of meritocracy writes as $[\rho^{\mathrm{m}}_{AA}, +\infty) \supset [\rho^{\mathrm{m}}, +\infty)$.

General proof for (i). Consider a representation threshold of $R = k - l$ with $l \in \{1, ..., k-2\}$. Then, denoting by $\tilde{V}$ the value function under affirmative action with representation threshold R, in any equilibrium

$$\begin{cases} \tilde{V}_{k+l} = \overline{x}s + \delta\left[\dfrac{k+l-1}{N-1}\tilde{V}_{k+l-1} + \dfrac{k-l}{N-1}\tilde{V}_{k+l}\right] \\ \tilde{V}_{k-l-1} = \overline{x}s + \delta\left[\dfrac{k-l-1}{N-1}\tilde{V}_{k-l-1} + \dfrac{k+l}{N-1}\tilde{V}_{k-l}\right] \end{cases}$$

Specializing to the meritocratic equilibrium (we omit superscripts for ease of notation),

$$\begin{cases} \tilde{u}_{k+l-1} = -xs - (1-x)b + \delta x\left[\dfrac{k-l}{N-1}\tilde{u}_{k+l-1} + \dfrac{k+l-2}{N-1}\tilde{u}_{k+l-2}\right] \\ \tilde{u}_{k-l-1} = xs - (1-x)b + \delta x\left[\dfrac{k-l-1}{N-1}\tilde{u}_{k-l-1} + \dfrac{k+l-1}{N-1}\tilde{u}_{k-l}\right] \end{cases}$$

We will first show that the sequence $(\tilde{u}_i)_{i \geq k-1}$ satisfies at least one of the following assertions: (A_1) it decreases with i, or (A_2) it is always strictly negative; and that in particular $\tilde{u}_{k+l-1} < 0$ [161]. As in the baseline case, the monotonicity property (A_1) implies that the most tempting deviation from meritocracy to entrenchment is when the majority has size k and faces an untalented ingroup candidate and a talented outgroup candidate, while if (A_2) holds, then all deviations to entrenchment are non-profitable as they yield a deviation payoff equal to

$$-(s-b) + \delta\left[\left(1 - \frac{i}{N-1}\right)\tilde{u}_i + \frac{i-1}{N-1}\tilde{u}_{i-1}\right] < 0$$

By contrast, the sign of $\tilde{u}_{k+l-1}$ suggests there may be profitable deviations from meritocracy with ties broken in favour of the majority candidate to meritocracy with ties broken in favour of the minority candidate when s/b is high enough. (Lastly, because of discounting, there can be no profitable deviation consisting in voting an untalented minority candidate instead of a talented majority one.)

We first suppose by contradiction that $\tilde{u}_{k+l-1} \geq 0$. The usual induction argument relying on (10) then yields that $\tilde{u}_{k-1} > \tilde{u}_k > ... > \tilde{u}_{k+l-1} \geq 0$. Yet, summing as in the proof of Lemma 1, the above recursive expression for $\tilde{u}_{k+l-1}$ with (13) and (10) over indices k to $k+l-2$, and rearranging, yields on

[161] A sketch of the proof is as follows: suppose by contradiction that $\tilde{u}_{k+l-1} \geq 0$, then by induction using (10), $\tilde{u}_{k-1} > ... > \tilde{u}_{k+l-1} \geq 0$. Yet summing the recursive expession of $\tilde{u}_i$ for $i \in \{k-1, ..., k+l-1\}$ implies that $\tilde{u}_{k-2} \geq 0$ as the sum of flow differential payoffs is equal to $-xs - (1-x)b + (1-2x)b = -x(s+b) < 0$. Using repeatedly the same argument gives that $u_i \geq 0$ for all $i \in \{k-l-1, ..., k+l-1\}$. A contradiction then obtains by summing the recursive expressions of u_i over the indices $i \in \{k-l-1, ..., k+l-1\}$, and noting that the sum of flow differential payoffs is equal to $-b < 0$.

the LHS a weighted sum of $\tilde{u}_{k-1}, ..., \tilde{u}_{k+l-1}$ which is strictly positive, while on the RHS:

$$-xs - (1-x)b + (1-2x)b + \delta\frac{k-2}{N-1}\tilde{u}_{k-2} = -x(s+b) + \delta(1-x)\frac{k-2}{N-1}\tilde{u}_{k-2},$$

and so $\tilde{u}_{k-2} > 0$. Summing (12) in $k-2$ to the above sum, and rearranging, yields on the LHS a weighted sum of $\tilde{u}_{k-1}, ..., \tilde{u}_{k+l-1}$ which is strictly positive, and on the RHS:

$$-x(s+b) + \delta(1-x)\frac{k-3}{N-1}\tilde{u}_{k-3},$$

Hence, $\tilde{u}_{k-3} > 0$, and by repeating this argument, $\tilde{u}_i > 0$ for any $i \in \{k-l-1, ..., k+l-1\}$. Yet summing the above recursive expressions of $\tilde{u}_{k-l-1}$ and $\tilde{u}_{k+l-1}$ together with (10)-(12)-(13) for $i \in \{k-l, ..., k+l-2\}$, yields after rearranging, on the LHS a weighted sum of all $\tilde{u}_i$ which is strictly positive, while on the RHS: $-x(s+b) + xs - (1-x)b = -b < 0$, which is a contradiction. Consequently, $\tilde{u}_{k+l-1} < 0$.

In order to show that the sequence $(\tilde{u}_i)_{i \geq k-1}$ satisfies either (A_1) or (A_2) (or both), we proceed by induction considering the lowest index i^- such that $\tilde{u}_i < 0$ for any $i \geq i^-$. We first note that if $i^- \geq k$, then (10) brings by induction that[162]

$$\tilde{u}_{k+l-1} < \tilde{u}_{k+l-2} < ... < \tilde{u}_{i^-+1} < \tilde{u}_{i^-} < 0 < \tilde{u}_{i^--1} < \tilde{u}_{i^--2} < ... < \tilde{u}_{k-1},$$

which yields that (A_1) holds. If $i^- \leq k-1$, then (A_2) holds. By contrast, in the baseline setting without affirmative action, the sequence $(u_i)_{i \geq k-1}$ is positive for any i and decreases with i.

Consequently, in order to show that the existence region of meritocracy expands for low s/b with affirmative action, we only need to consider deviations from meritocracy to entrenchment when the majority is tight and faces an untalented ingroup candidate and a talented outgroup one, and show that the condition for non-profitability is looser with affirmative action than in the baseline setting.

Explicit computations yield

$$\begin{cases} \tilde{u}_{k+l-1} = -xs - (1-x)b + \delta x\left[\dfrac{k-l}{N-1}\tilde{u}_{k+l-1} + \dfrac{k+l-2}{N-1}\tilde{u}_{k+l-2}\right] \\ \tilde{u}_{k-l-1} = xs - (1-x)b + \delta x\left[\dfrac{k-l-1}{N-1}\tilde{u}_{k-l-1} + \dfrac{k+l-1}{N-1}\tilde{u}_{k-l}\right] \end{cases}$$

Thus using (10) in $k+l-1$ and (12) in $k-l-1$, together with the fact that $u_i \geq 0$ for all i in the baseline setting, one gets[163]

$$\begin{cases} \left[1 - \delta x\dfrac{k-l}{N-1}\right]\left(\tilde{u}_{k+l-1} - u_{k+l-1}\right) < -xs - (1-x)b + \delta x\dfrac{k+l-2}{N-1}\left(\tilde{u}_{k+l-2} - u_{k+l-2}\right) \\ \left[1 - \delta x\dfrac{k-l-1}{N-1}\right]\left(\tilde{u}_{k-l-1} - u_{k-l-1}\right) < xs - (1-x)b + \delta x\dfrac{k+l-2}{N-1}\left(\tilde{u}_{k-l} - u_{k-l}\right) \end{cases}$$

[162] The inequalities $\tilde{u}_{k+l-1} < \tilde{u}_{k+l-2} < ... < \tilde{u}_{i^-+1} < \tilde{u}_{i^-} < 0$ can be established by induction using the recursive expressions of the $\tilde{u}_i$ from $i = i^-$ up to $i = k+l-2$.

[163] Note that the omitted terms write for the first equation as

$$-\delta(1-x)\left[\frac{k+l-1}{N-1}u_{k+l-1} + \frac{k-l-1}{N-1}u_{k+l}\right],$$

which is thus proportional to $(-b)$ (see proof of Lemma 1 for details). Similarly for the second equation.

Therefore, using (10) in $k+l-2$ and (12) in $k-l$, one gets

$$\begin{cases} \left[1-\delta x\dfrac{k-l+1}{N-1}-\delta(1-x)\dfrac{k+l-2}{N-1}-\delta(1-x)\dfrac{k-l}{N-1}\dfrac{\delta x\dfrac{k+l-2}{N-1}}{1-\delta x\dfrac{k-l}{N-1}}\right]\big(\tilde{u}_{k+l-2}-u_{k+l-2}\big) \\ \qquad < \dfrac{\delta(1-x)\dfrac{k-l}{N-1}}{1-\delta x\dfrac{k-l}{N-1}}\big[-xs-(1-x)b\big]+\delta x\dfrac{k+l-3}{N-1}\big(\tilde{u}_{k+l-3}-u_{k+l-3}\big) \\ \left[1-\delta x\dfrac{k-l}{N-1}-\delta(1-x)\dfrac{k+l-1}{N-1}-\delta(1-x)\dfrac{k-l-1}{N-1}\dfrac{\delta x\dfrac{k+l-1}{N-1}}{1-\delta x\dfrac{k-l-1}{N-1}}\right]\big(\tilde{u}_{k-l}-u_{k-l}\big) \\ \qquad < \dfrac{\delta(1-x)\dfrac{k-l-1}{N-1}}{1-\delta x\dfrac{k-l-1}{N-1}}\big[xs-(1-x)b\big]+\delta x\dfrac{k+l-2}{N-1}\big(\tilde{u}_{k-l+1}-u_{k-l+1}\big) \end{cases}$$

We begin by noting that

$$\frac{k-l}{N-1}\left[1-\delta x\frac{k-l-1}{N-1}\right]>\frac{k-l-1}{N-1}\left[1-\delta x\frac{k-l}{N-1}\right],\qquad\text{and}\qquad\frac{\delta(1-x)\dfrac{k-l}{N-1}}{1-\delta x\dfrac{k-l}{N-1}}>\frac{\delta(1-x)\dfrac{k-l-1}{N-1}}{1-\delta x\dfrac{k-l-1}{N-1}}$$

We then observe that: $(k-l)(k+l-2)=(k-l+1)(k+l-1)-(2k-1)$, and as a consequence, using the above inequality,

$$\begin{aligned}&\left(\frac{k-l}{N-1}\right)^2\frac{k+l-2}{N-1}\left[1-\delta x\frac{k-l-1}{N-1}\right]\\ &\qquad>\frac{k-l+1}{N-1}\frac{k-l-1}{N-1}\frac{k+l-1}{N-1}\left[1-\delta x\frac{k-l}{N-1}\right]-\frac{k-l}{N-1}\left[1-\delta x\frac{k-l-1}{N-1}\right]\frac{1}{N-1},\end{aligned}$$

Using the fact that $\delta(1-x)\dfrac{k-l}{N-1}<1-\delta x\dfrac{k-l}{N-1}$, we get that

$$\begin{aligned}&\delta x\delta(1-x)\left(\frac{k-l}{N-1}\right)^2\frac{k+l-2}{N-1}\left[1-\delta x\frac{k-l-1}{N-1}\right]\\ &\qquad>\delta x\delta(1-x)\frac{k-l+1}{N-1}\frac{k-l-1}{N-1}\frac{k+l-1}{N-1}\left[1-\delta x\frac{k-l}{N-1}\right]-\frac{\delta x}{N-1},\end{aligned}$$

Hence, since

$$\frac{k-l+1}{N-1}\left[1-\delta(1-x)\frac{k+l-1}{N-1}\right]=\frac{k-l}{N-1}\left[1-\delta(1-x)\frac{k+l-2}{N-1}\right]+\frac{1-\delta(1-x)}{N-1}$$

we get that,

$$\frac{k-l+1}{N-1}\left[1-\delta x\frac{k-l}{N-1}-\delta(1-x)\frac{k+l-1}{N-1}-\delta(1-x)\frac{k-l-1}{N-1}\frac{\delta x\frac{k+l-1}{N-1}}{1-\delta x\frac{k-l-1}{N-1}}\right]$$
$$>\frac{k-l}{N-1}\left[1-\delta x\frac{k-l+1}{N-1}-\delta(1-x)\frac{k+l-2}{N-1}-\delta(1-x)\frac{k-l}{N-1}\frac{\delta x\frac{k+l-2}{N-1}}{1-\delta x\frac{k-l}{N-1}}\right]$$

By downward (resp. upward) induction on $(\tilde{u}_i-u_i)$ for $i\geq k$ (resp. for $i\leq k-2$), we will get that

$$C_1\big(\tilde{u}_{k-1}-u_{k-1}\big)<-C_2xs-C_3(1-x)b<0$$

where C_1, C_2 and C_3 are strictly positive constants that depend on the parameters k, l and x. We sketch the induction argument. Using (10)-(12), we obtain two sequences $(a_j)_{0\leq j\leq l-2}$ and $(b_j)_{0\leq j\leq l-2}$ such that for any $j\leq l-2$,

$$\begin{cases} a_j\big(\tilde{u}_{k+j}-u_{k+j}\big)<-\big[xs+(1-x)b\big]\dfrac{\delta(1-x)\frac{k-l}{N-1}}{1-\delta x\frac{k-l}{N-1}}\displaystyle\prod_{n=j+1}^{l-2}\left(\frac{\delta(1-x)}{a_n}\frac{k-n-1}{N-1}\right)+\delta x\frac{k+j-1}{N-1}\big(\tilde{u}_{k+j-1}-u_{k+j-1}\big) \\ b_j\big(\tilde{u}_{k-j-2}-u_{k-j-2}\big)<-\big[xs+(1-x)b\big]\dfrac{\delta(1-x)\frac{k-l-1}{N-1}}{1-\delta x\frac{k-l-1}{N-1}}\displaystyle\prod_{n=j+1}^{l-2}\left(\frac{\delta(1-x)}{b_n}\frac{k-n-2}{N-1}\right)+\delta x\frac{k+j}{N-1}\big(\tilde{u}_{k-j-1}-u_{k-j-1}\big) \end{cases}$$

where

$$\begin{cases} a_{j-1}=1-\delta x\dfrac{k-j}{N-1}-\delta(1-x)\dfrac{k+j-1}{N-1}-\delta(1-x)\dfrac{k-j-1}{N-1}\dfrac{\delta x\frac{k+j-1}{N-1}}{a_j} \\ b_{j-1}=1-\delta x\dfrac{k-j-1}{N-1}-\delta(1-x)\dfrac{k+j}{N-1}-\delta(1-x)\dfrac{k-j-2}{N-1}\dfrac{\delta x\frac{k+j}{N-1}}{b_j} \end{cases}$$

We first note that by induction[164]

$$\forall j\leq l-1,\qquad \frac{\delta(1-x)}{a_j}\frac{k-j-1}{N-1}<1 \tag{57}$$

We then show by downward induction on j that for any $j\leq l-2$,

$$\frac{1}{a_j}\frac{k-j-1}{N-1}>\frac{1}{b_j}\frac{k-j-2}{N-1},$$

which yields the result. The initialization ($j=l-2$) has been established above. As for the induction,

[164]The initialization with $j=l-1$ stems from the observation that

$$\delta(1-x)\frac{k-l}{N-1}<1-\delta x\frac{k-l}{N-1}$$

i.e. to show that $a_{j-1}(k-j-1) < b_{j-1}(k-j)$, one notes that for any $j \geq 0$,

$$\frac{k-j}{N-1}\left[1-\delta(1-x)\frac{k+j}{N-1}\right] = \frac{k-j-1}{N-1}\left[1-\delta(1-x)\frac{k+j-1}{N-1}\right] + \frac{1-\delta(1-x)}{N-1}$$

and $(k-j-1)(k+j-1) = (k-j)(k+j) - (2k-1)$, which implies with the induction hypothesis that

$$\frac{k-j-1}{N-1}\frac{k+j-1}{N-1}\frac{1}{a_j}\frac{k-j-1}{N-1} > \frac{k-j}{N-1}\frac{k+j}{N-1}\frac{1}{b_j}\frac{k-j-2}{N-1} - \frac{1}{a_j}\frac{k-j-1}{N-1}\frac{1}{N-1}$$

and thus, using (57),

$$\delta x\delta(1-x)\frac{k-j-1}{N-1}\frac{k+j-1}{N-1}\frac{1}{a_j}\frac{k-j-1}{N-1} > \delta x\delta(1-x)\frac{k-j}{N-1}\frac{k+j}{N-1}\frac{1}{b_j}\frac{k-j-2}{N-1} - \frac{\delta x}{N-1}$$

Therefore, using the recursive expression of a_{j-1} and b_{j-1}, we have that

$$a_{j-1}(k-j-1) < b_{j-1}(k-j) - \frac{1-\delta}{N-1} < b_{j-1}(k-j),$$

as was to be shown.

This in turn implies that $\left(\tilde{u}_k - u_k\right) < 0$. Therefore, the non-profitability conditions for deviations from meritocracy to entrenchment is (strictly) looser with a representation threshold R than without. Moreover, since C_2 and C_3 are strictly positive, and since the omitted negative terms are all proportional to $(-b)$ (see footnote 163), the existence region of meritocracy expands downard (i.e. for low values of s/b).[165]

For s/b sufficiently high, there exists a unique equilibrium: the majority always picks the most talented candidate and breaks ties in favour of the minority candidate. By the above argument, this equilibrium exists for s/b sufficiently high[166]. We now show it is unique for $b = 0 < s$ among meritocratic equilibria. The above computations apply: indeed, the argument follows by letting $b = 0 < s$, noting that in the unconstrained meritocratic equilibrium, this implies $u_i = 0$ for any $i \in \{1, ..., N-2\}$, and thus with the above argument $\tilde{u}_i \leq 0$ for any $i \geq k-1$. Hence, the deviation differential payoff from standard-favoritism meritocracy to reverse-favoritism meritocracy at majority size i is given by

$$-\left(\frac{k-1+i}{N-1}\tilde{u}_i^{\mathrm{m}} + \frac{k-1-i}{N-1}\tilde{u}_{i-1}^{\mathrm{m}}\right) > 0,$$

which yields the result.

An intuition underlying this result is that, whenever there is a vertical tie, if the majority recruits its own candidate, it increases the chances of reaching the representation threshold which is costly in terms of quality. In contrast, by recruiting the minority's candidate, the majority loses no homophily benefit (since $b = 0$), while lowering the chances of reaching the representation threshold. Hence, fixing s, since

[165] Moreover, since either (A_1) or (A_2) hold, we have by monotonicity with respect to b and s that for any value of the ratio s/b such that meritocracy exists, then for any higher value of the ratio, there can be no profitable deviations towards entrenchment (i.e. un-meritocratic decisions are always unprofitable). This establishes that meritocratic decisions – with ties broken either in favour of the majority or the minority – happen on a half-line, i.e. for any s/b sufficiently high.

[166] This can be shown by letting $b = 0 < s$ and noting that in the reverse-favoritism meritocratic equilibrium, this implies that $u_i \leq 0$ for any $i \geq k-1$. Therefore, there exists no profitable deviations (in particular when both candidates are equally talented.

value functions are continuous with respect to b, the existence and uniqueness of this equilibrium holds for b in a neighbourhood of 0. Therefore, the result obtains for any s/b sufficiently high.

O.2 Proof of Proposition 13-(ii)

Let $N \geq 4$ and $1 \leq l \leq k-1$. The ergodic aggregate efficiency of a canonically entrenched organization under laissez-faire and a meritocratic one under affirmative action with representation threshold l write respectively:

$$\begin{cases} S^{\mathrm{e}} = N(N-1)\left[\dfrac{k+1}{N}\nu^{\mathrm{e}}_{k+1}\overline{x} + \left(1 - \dfrac{k+1}{N}\nu^{\mathrm{e}}_{k+1}\right)(\overline{x}+x)\right]\tilde{s} \\ S^{\mathrm{m,AA}} = N(N-1)\left[\dfrac{l}{N}\nu^{\mathrm{m,AA}}_{N-l}\overline{x} + \left(1 - \dfrac{l}{N}\nu^{\mathrm{m,AA}}_{N-l}\right)(\overline{x}+x)\right]\tilde{s} \end{cases}$$

and thus:

$$S^{\mathrm{m,AA}} - S^{\mathrm{e}} = N(N-1)\left[\frac{k+1}{N}\nu^{\mathrm{e}}_{k+1} - \frac{l}{N}\nu^{\mathrm{m,AA}}_{N-l}\right]x\tilde{s}$$

Explicit computations (see Lemma 2 and its proof in Section E) yield:

$$\begin{cases} \nu^{\mathrm{e}}_{k+1}\left[1 + \displaystyle\sum_{i=1}^{k-1}\left(\frac{1-x}{x}\right)^i \prod_{j=1}^{i}\frac{k-j}{k+1+j}\right] = 1 \\ \nu^{\mathrm{m,AA}}_{N-l}\left[1 + \displaystyle\sum_{i=1}^{k-l-1}\left(\frac{x}{1-x}\right)^i \prod_{j=1}^{i}\frac{N-l+1-j}{l+j} + \left(\frac{x}{1-x}\right)^{k-l}\frac{k+1}{N}\prod_{j=1}^{k-l-1}\frac{N-l+1-j}{l+j}\right] = 1 \end{cases}$$

Consequently, $S^{\mathrm{m,AA}} - S^{\mathrm{e}}$ has same sign as

$$\begin{aligned}(k+1)&\left[1 + \sum_{i=1}^{k-l-1}\left(\frac{x}{1-x}\right)^i \prod_{j=1}^{i}\frac{N-l+1-j}{l+j} + \left(\frac{x}{1-x}\right)^{k-l}\frac{k+1}{N}\prod_{j=1}^{k-l-1}\frac{N-l+1-j}{l+j}\right] \\ &- l\left[1 + \sum_{i=1}^{k-1}\left(\frac{1-x}{x}\right)^i \prod_{j=1}^{i}\frac{k-j}{k+1+j}\right]\end{aligned}$$

We then note that the above expression is strictly negative for x in a neighbourhood of 0, and strictly positive for x in a neighbourhood of 1. Moreover, since $x/(1-x)$ (resp. $(1-x)/x$) strictly increases (resp. decreases) with $x \in (0, 1/2)$, there exists a unique $x_{\mathrm{AA}}(l) \in (0, 1/2]$ such that for any $x < x_{\mathrm{AA}}(l)$ (resp. $x > x_{\mathrm{AA}}(l)$), the above expression is strictly negative (resp. positive).

Lastly, we note that by construction, $x_{\mathrm{AA}}(l)$ is such that

$$\begin{aligned}(k+1)&\left[1 + \sum_{i=1}^{k-l-1}\left(\frac{x_{\mathrm{AA}}(l)}{1-x_{\mathrm{AA}}(l)}\right)^i \prod_{j=1}^{i}\frac{N-l+1-j}{l+j} + \left(\frac{x_{\mathrm{AA}}(l)}{1-x_{\mathrm{AA}}(l)}\right)^{k-l}\frac{k+1}{N}\prod_{j=1}^{k-l-1}\frac{N-l+1-j}{l+j}\right] \\ &= l\left[1 + \sum_{i=1}^{k-1}\left(\frac{1-x_{\mathrm{AA}}(l)}{x_{\mathrm{AA}}(l)}\right)^i \prod_{j=1}^{i}\frac{k-j}{k+1+j}\right]\end{aligned}$$

The LHS in the above equation strictly decreases with l for any given x fixed, and strictly increases with x for any fixed l. By contrast, the RHS strictly increases with l for any fixed x, and strictly decreases

with x for any fixed l. Hence $x_{\mathrm{AA}}(l)$ strictly increases with l.

P Supermajority electoral rules

We assume that the principal does not observe the candidates' talent. Hence if no candidate reaches the election threshold, the principal picks one among the two at random. Consequently, the principal's blindness makes failing to reach the election threshold costly for majority members.

With a majority+l voting rule, whenever the majority has size $M \geq k + l$, the principal does not intervene. By contrast, for any lower majority size, minority members are pivotal. Consequently, two opposite effects drive the results:

- The principal's blind intervention if the supermajority is not reached may make meritocracy relatively more attractive and prevent the organization from being entrenched.
- Super-entrenchment at level l shields the entrenched majority from the principal's intervention.

We define the "meritocratic equilibrium" as an equilibrium in which each group votes for its own candidate if and only if she is at least as talented as the rival candidate. We say the minority is entrenched if, whenever it is sufficiently large so as to be able to block nominations, it always votes for its own candidate, and define the "level-l entrenched equilibrium" as the equilibrium in which the minority is entrenched and the majority super-entrenched at level l. We look for monotonic symmetric MPEs in weakly undominated strategies, which indeed exist.

Proposition 15. ***(Supermajority electoral rules)*** *Let $x < 1/2$. Let $l \geq 1$ and consider the majority+l voting rule.*

(i) For s/b sufficiently close to 1, super-entrenchment at level l is the unique symmetric MPE in weakly undominated strategies such that a stronger majority makes (weakly) more meritocratic recruitments.

(ii) For δ sufficiently low, the existence region of meritocracy widens with respect to laissez-faire.

As a by-product of the proof, we show that for any supermajority rule with parameter l (with $l = 0$ corresponding to the baseline model), (a) meritocracy exists if and only if there is no one-shot profitable deviation for the majority whenever it has size $k + l$ and faces a talented minority candidate and an untalented majority one (in other words, as intuitive, this is the most tempting situation for a group to deviate from meritocracy); and (b) for a low δ, the existence region of meritocracy is minimal for $l = 0$, i.e. with the simple majority rule.

We first prove that, for δ close to 0, the existence region of meritocracy widens (step (a)), before establishing the uniqueness within the class of monotonic symmetric MPEs, and existence of the level-l super-entrenchment equilibrium in a neighbourhood of 1 (steps (b) and (c)). We stress that we only show claim (a) for δ in a neighbourhood of 0, while we establish claims (b) and (c) for any discount factor $\delta \in [0, (N-1)/N)$.

(a) Consider the meritocratic strategy for both groups. Whenever the majority size is (weakly above) $k + l$ – or equivalently the minority size is (weakly) below $(N - k - l - 1)$ –, the majority picks the most talented candidate, breaking ties in favor of its own candidate. For any majority size $M \in \{k, ..., k+l-1\}$,

- if candidates' abilities differ, the most talented candidate is recruited,
- if both candidates are equally talented, the recruited candidate is drawn at random among the two (with equal probability).

Hence, for any majority size $M \in \{k, ..., k+l-1\}$, the average quality of a candidate is equal to $(\overline{x}+x)\tilde{s}$ as in the baseline meritocracy, yet the average homophily payoff accruing to majority and minority members is equal to $\tilde{b}/2$ instead of respectively $(1-x)\tilde{b}$ and $x\tilde{b}$.

For any $i \in \{1, ..., N-2\}$, let $\tilde{u}_i^{\mathrm{m}} \equiv \tilde{V}_{i+1}^{\mathrm{m}} - \tilde{V}_i^{\mathrm{m}}$ where $\tilde{V}_i^{\mathrm{m}}$ is the value function at group size i in the meritocratic equilibrium. By the usual argument (e.g. cf. proof of Proposition 1), the existence region of meritocracy strictly widens with respect to *laissez-faire* if and only if the meritocracy-to-entrenchment deviation differential payoffs are strictly lower than the highest such differential payoff under *laissez-faire*, i.e. if and only if for any $i \in \{k-l, ..., N-1\}$,

$$\tilde{\Delta}_i \equiv \frac{i-1}{N-1}\tilde{u}_{i-1}^{\mathrm{m}} + \left(1 - \frac{i}{N-1}\right)\tilde{u}_i^{\mathrm{m}} < \frac{k-1}{N-1}\left(u_{k-1}^{\mathrm{m}} + u_k^{\mathrm{m}}\right) \equiv \Delta_k$$

The result thus clearly holds for δ in a neighbourhood of 0 [167]. We assume in the following that $\delta > 0$. We establish a few additional facts in favour of our conjecture that the result holds for any $\delta \in [0, (N-1)/N)$.

Namely, the sequence $(u_i^{\mathrm{m}})_i$ is given by Equations (10)-(12)-(13), while the sequence $(\tilde{u}_i^{\mathrm{m}})_i$ satisfies (10) for any $i \geq k+l$, and (12) for any $i \leq k-l-2$. By contrast, for any $i \in \{k-l-1, ..., k+l-1\}$, $\tilde{u}_i^{\mathrm{m}}$ is given by[168]

$$\begin{cases} \tilde{u}_{k+l-1}^{\mathrm{m}} = \dfrac{1-2x}{2}b + \delta(1-x)\left[\dfrac{k+l-1}{N-1}\tilde{u}_{k+l-1}^{\mathrm{m}} + \left(1 - \dfrac{k+l}{N-1}\right)\tilde{u}_{k+l}^{\mathrm{m}}\right] \\ \qquad\qquad + \dfrac{\delta}{2}\left[\dfrac{k+l-2}{N-1}\tilde{u}_{k+l-2}^{\mathrm{m}} + \left(1 - \dfrac{k+l-1}{N-1}\right)\tilde{u}_{k+l-1}^{\mathrm{m}}\right] \\ \tilde{u}_i^{\mathrm{m}} = \dfrac{\delta}{2}\left[\dfrac{i-1}{N-1}\tilde{u}_{i-1}^{\mathrm{m}} + \left(1 - \dfrac{i}{N-1}\right)\tilde{u}_i^{\mathrm{m}}\right] + \dfrac{\delta}{2}\left[\dfrac{i}{N-1}\tilde{u}_i^{\mathrm{m}} + \left(1 - \dfrac{i+1}{N-1}\right)\tilde{u}_{i+1}^{\mathrm{m}}\right], \quad \forall i \in \{k-l, ..., k+l-2\} \\ \tilde{u}_{k-l-1}^{\mathrm{m}} = \dfrac{1-2x}{2}b + \delta(1-x)\left[\dfrac{k-l-2}{N-1}\tilde{u}_{k-l-2}^{\mathrm{m}} + \left(1 - \dfrac{k-l-1}{N-1}\right)\tilde{u}_{k-l-1}^{\mathrm{m}}\right] \\ \qquad\qquad + \dfrac{\delta}{2}\left[\dfrac{k-l-1}{N-1}\tilde{u}_{k-l-1}^{\mathrm{m}} + \left(1 - \dfrac{k-l}{N-1}\right)\tilde{u}_{k-l}^{\mathrm{m}}\right] \end{cases} \tag{58}$$

The usual arguments (see below the proof of claim (b)) yield that for any discount factor, $0 < \tilde{u}_1^{\mathrm{m}} < ... < \tilde{u}_{k-l-1}^{\mathrm{m}}$ and $\tilde{u}_{k+l-1}^{\mathrm{m}} > ... > \tilde{u}_{N-2}^{\mathrm{m}} > 0$ [169]. Moreover, (58) implies that the sequence $(\tilde{u}_i^{\mathrm{m}})_i$ does not reach its global maximum for $i \in \{k-l, ..., k+l-2\}$. Hence the maximum of the sequence $(\tilde{u}_i^{\mathrm{m}})_i$ is either

[167] For $\delta = 0$, one has $u_{k-1}^{\mathrm{m}} = (1-2x)b = 2\tilde{u}_{k-l-1} = 2\tilde{u}_{k+l-1}$, while $u_i = 0$ for any $i \neq k-1$, and $\tilde{u}_i = 0$ for any $i \notin \{k-l-1, k+l-1\}$. Therefore,

$$\max_{i \in \{k-l, ..., N-1\}} \tilde{\Delta}_i = \tilde{\Delta}_{k+l} = \frac{k+l-1}{N-1}\frac{1-2x}{2}b \leq \frac{k-1}{N-1}(1-2x)b$$

[168] The system derives from the explicit expression of the value functions after rearranging.

[169] Note that this implies that

$$\tilde{\Delta}_i = \frac{i-1}{N-1}\tilde{u}_{i-1}^{\mathrm{m}} + \left(1 - \frac{i}{N-1}\right)\tilde{u}_i^{\mathrm{m}}$$

increases (resp. decreases) with $i \in \{1, ..., k-l-1\}$ (resp. $i \in \{k+l-1, ..., N-2\}$), which implies that an equivalence condition for our result needs only consider deviation differential payoffs at sizes $i \in \{k-l, ..., k+l\}$ instead of the larger set of sizes $\{k-l, ..., N-1\}$.

$\tilde{u}^{\mathrm{m}}_{k-l-1}$ or $\tilde{u}^{\mathrm{m}}_{k+l-1}$.

Similarly, by definition of $\tilde{\Delta}$ and by rearranging, for any i,

$$\tilde{\Delta}_{i+1} - \tilde{\Delta}_i = \frac{i-1}{N-1}(\tilde{u}^{\mathrm{m}}_i - \tilde{u}^{\mathrm{m}}_{i-1}) + \left(1 - \frac{i+1}{N-1}\right)(\tilde{u}^{\mathrm{m}}_{i+1} - \tilde{u}^{\mathrm{m}}_i),$$

and thus the sequence $(\tilde{\Delta}_i)_i$ inherits the monotonicity (if any) of the sequence $(\tilde{u}_i)_i$. Moreover, (58) implies that for any $i \in \{k-l+1, ..., k+l-2\}$,

$$\tilde{\Delta}_i = \frac{\delta}{2}\left[\frac{i-1}{N-1}\big(\tilde{\Delta}_{i-1} + \tilde{\Delta}_i\big) + \left(1 - \frac{i}{N-1}\right)\big(\tilde{\Delta}_i + \tilde{\Delta}_{i+1}\big)\right], \tag{59}$$

and thus the sequence $(\tilde{\Delta}_i)_i$ does not reach its global maximum over $\{k-l+1, ..., k+l-2\}$ (namely, the sequence $(\tilde{\Delta}_i)_i$ is either monotonic or U-shaped on $\{k-l, ..., k+l-1\}$).. Therefore the sequence reaches its maximum in one of the elements of the set $\{k-l-1, k-l, k+l-1, k+l\}$. Hence, the existence region of meritocracy strictly widens with respect to *laissez-faire* if and only if [170]

$$\max\left\{\tilde{\Delta}_{k-l}, \tilde{\Delta}_{k+l-1}, \tilde{\Delta}_{k+l}\right\} \leq \Delta_k$$

By construction, for $\delta = 0$, we have that for any $l \in \{1..., k-1\}$,

$$\tilde{\Delta}_{k+l} = \frac{k+l-1}{N-1}\frac{1-2x}{2}b, \qquad \tilde{\Delta}_{k+l-1} = \frac{k-l}{N-1}\frac{1-2x}{2}b, \qquad \tilde{\Delta}_{k-l} = \frac{k-l-1}{N-1}\frac{1-2x}{2}b,$$
$$\tilde{\Delta}_{k-l-1} = \frac{k+l}{N-1}\frac{1-2x}{2}b, \qquad \text{and} \qquad \tilde{\Delta}_i = 0 \quad \text{for any } i \notin \{k-l-1, k-l, k+l-1, k+l\},$$

while

$$\Delta_k = \frac{N-2}{N-1}\frac{1-2x}{2}b, \qquad \Delta_{k-1} = \frac{N}{N-1}\frac{1-2x}{2}b, \qquad \text{and} \qquad \Delta_i = 0 \quad \text{for any } i \notin \{k-1, k\}$$

Hence, the above discussion implies that for δ small, the existence region of meritocracy strictly widens with respect to *laissez-faire* for any $l \in \{1, ..., k-2\}$. Lastly, a first-order Taylor expansion for δ small in the case $l = k-1$ yields[171]

$$\begin{cases} \tilde{\Delta}_{N-1} = \dfrac{k-1}{N-1}(1-2x)b + \delta(1-2x)b\dfrac{k-1}{N-1}\left[\dfrac{1}{2(N-1)} + (1-x)\dfrac{N-2}{N-1}\right] + O(\delta^2) \\ \Delta_k = \dfrac{k-1}{N-1}(1-2x)b + \delta(1-2x)b\dfrac{k-1}{N-1}\left[x\dfrac{k-1}{N-1} + 1 - x\right] + O(\delta^2) \end{cases}$$

and thus, $\tilde{\Delta}_{N-1} < \Delta_k$. In other terms, for δ close to 0, the existence region of meritocracy strictly widens under the unanimity rule with respect to *laissez-faire*.

(b) The uniqueness of the level-l super-entrenchment equilibrium within the class of symmetric MPEs

[170] Recall that $\tilde{\Delta}_{k-l-1}$ plays no role since the minority has no control when it has size $k-l-1$. Moreover, one can show that for any discount factor,

$$\max\left\{\tilde{\Delta}_{k-l}, \tilde{\Delta}_{k+l-1}, \tilde{\Delta}_{k+l}\right\} = \tilde{\Delta}_{k+l}$$

[171] Indeed, for δ in a neighbourhood of 0, $\max\left\{\tilde{\Delta}_{k-l}, \tilde{\Delta}_{k+l-1}, \tilde{\Delta}_{k+l}\right\} = \tilde{\Delta}_{k+l}$.

such that a stronger majority makes more meritocratic recruitments for s/b in a neighbourhood of 1 derives from the argument in the proof of Proposition 6 (see Section I) with the probabilities of the majority losing the vote being equal to $\Lambda(M) = 1/2$ for $M \in \{k, ..., k+l-1\}$ and to $\Lambda(M) = 0$ for any $M \geq k+l\}$.

(c) We resort to the usual argument for the existence of the level-l super-entrenchment equilibrium. Consider $s = b > 0$, and let the superscript "e+" denote super-entrenchment at level l. The deviation differential payoff from super-entrenchment at level l to a lower level of super-entrenchment in $i \in \{k-l, ..., k+l\}$ writes as

$$-\frac{i-1}{N-1}\tilde{u}_{i-1}^{e+} - \left(1 - \frac{i}{N-1}\right)\tilde{u}_i^{e+} \leq 0,$$

where the inequality derives from standard computations which yield that $\tilde{u}_i^{e+} \geq 0$ for any $i \geq k-1$ (where the inequality is strict whenever $i \in \{k-1, ..., k+l-1\}$ and with $u_i = 0$ for any $i \geq k+l$, see above and proof of Lemma 1). Hence, by continuity, super-entrenchment at level l is an equilibrium for s/b in a neighbourhood of 1.