Meritocracy and Homophily in Collegial Organizations

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

October 2, 2024

Abstract

In collegial organizations, homophily jeopardizes the recruitment of the most talented ("meritocracy"). We analyze the dynamics and welfare properties of an organization whose incumbent members coopt new ones in a forward-looking manner. We identify organizations in which meritocracy is likely to give way to favoritism and entrenchment, and investigate policy interventions (such as affirmative action, quality assessment exercises, overruling of majority decisions) and their unintended consequences.

Keywords: Cooptation, organizations, Markov games, meritocracy, homophily, affirmative action, glass ceiling, 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 numerous seminars and conferences for helpful suggestions. Funding from the European Research Council (grant agreement no. 101098319 - ERC HWS) is gratefully acknowledged.

[*]Paris School of Economics and Ecole des Ponts ParisTech, paul-henri.moisson@psemail.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. Most often, new members are coopted,[1] i.e., recruited by incumbent members according to a (formal or informal) voting process, perhaps under constraints imposed by internal rules or external intervention. Even in more hierarchical organizations, recruitment often reflects some degree of collegiality, whereby whoever holds the formal authority on the recruitment decision puts substantial weight on subordinates' (candidates' would-be colleagues) opinions. We aim at studying the consequences of such collegiality.

This paper analyzes the Markovian dynamics, the discrimination in hiring (and promotion), and the welfare properties of an organization whose members are forward-looking and are driven by two motives in their cooptation decisions: talent and homophily. All else being equal, all members prefer a more talented candidate to a less talented one. However, homophily along a "horizontal" trait (gender, religion, ethnicity, politics, scientific field or approach, values, family, friendship, class loyalty...) makes members prefer an in-group candidate to an out-group one unless the latter is substantially more talented than the former. This misalignment of horizontal preferences creates a benefit from controlling the organization's recruitments. Members of the majority group may favor an in-group candidate over a more talented out-group one, so as to stuff the organization with their chums and thereby entrench their grip on the organization's recruitments. This violates meritocracy, defined here as the selection of the more talented.

The first contribution of our paper is to derive theoretical predictions about the drivers of entrenchment in the absence of public intervention. *(a) Homophilic intensity.* An obvious factor is the preference for homophily: agents who attach more importance to mingling with their in-group are more likely to give up on meritocracy. *(b) Weak-link principle.* It takes two to build meritocracy: the cooptation process is only as meritocratic as the group with the strongest homophily preferences allows it. *(c) Uncertainty about future control.* A thinner majority is more prone to depart from meritocratic hiring. The larger the majority

[1] We focus on "cooptation" in the sense of "periodic selection by incumbent members of new members, according to a given voting rule". 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). Introducing the possibility that coopting outsiders may change their behaviour and safeguard the organization is a straightforward yet interesting extension of our model – for instance, the outsiders' nuisance power could be captured by assuming that the probability that the organization continues falls sharply when it is too monolithic (e.g., due to the prospect of a "revolution").

size, the tighter the majority's grip on the organization, the milder the discrimination against minority candidates. Moreover, when voting participation is uncertain or when allegiance to the in-group is assessed with noise, the majority may optimally build a buffer against unexpected losses of control. This "buffer effect" will similarly drive some unintended consequences of policy interventions (see below). *(d) Homogamic evaluation capability.* Talent evaluation that is more accurate within the in-group (due to field expertise or familiarity) under weak conditions raises entrenchment, as the majority is less able to identify talent within the pool of minority candidates. *(e) Patience.* When facing a trade-off between coopting a minority member or a less talented majority one, the majority members weigh the discounted benefits from the talent differential against the future control benefits. We show that the former have a relatively lower weight when agents have a more sustained relationship with the organization: a longer expected tenure (a longer time-horizon) within the organization makes majority members more prone to entrench the organization. *(f) Availability duration.* When rejected candidates are likely to re-apply, the cost for the majority to secure its entrenchment by turning down a talented minority candidate decreases, making meritocracy violations more attractive to the majority.

As allowed by our framework, entrenchment is not always socially detrimental – for instance, friendship circles are often based on homophily in tastes. We emphasize a fundamental asymmetry between entrenchment and meritocracy, though: While entrenchment always prevails whenever it is socially desirable, there exists a range of parameter values such that meritocracy is desirable and yet is violated in equilibrium. We refer to the cause of this discrepancy as a *collegial bias against meritocracy*, which arises from talent being scarce and strength lying in numbers.

Our second contribution is an analysis of familiar public policy interventions to promote meritocracy. This paper provides an analysis (the first to the best of our knowledge) of three common policy interventions on a collegial organization: the overruling of majority decisions, quality assessment exercises and affirmative action. These interventions differ not only in the information required to set them in motion, but also in their efficacy at achieving their meritocratic objective.

Firstly, the *occasional overruling of hiring decisions* can backfire by increasing the majority's incentive to entrench, as the majority builds a larger buffer to reduce the probability of its losing control. The bottom line is that unless they take continuous control over the recruitment, even well-intentioned and well-informed public officials may well do more harm than good. Secondly, *financial rewards for quality* (for example through research assessment exercises and research councils' grants) promote meritocracy, but they must be targeted to where they have the most impact, that is organizations that under laissez-faire are neither strongly entrenched nor naturally meritocratic. Thirdly, *affirmative action* (in the form of a minority quota) can switch the organization's regime from entrenchment to meritocracy by reducing the value for majority members of controlling recruitments. The policy however is costly when the meritocratic decision would select the majority candidate, but the minority quota is binding, forcing the recruitment of the minority candidate. Overall, well-meaning policies cannot be presumed to raise welfare without considering the organizational response they trigger.

Technical contribution and roadmap.

Section 2 builds the baseline model under laissez-faire. As we later show, this model gives the best chance to meritocracy by assuming that (i) the majority can perfectly predict prospective hires' allegiance and there is no uncertainty about turnout in the recruitment elections, and (ii) groups can identify talent equally well for out-group candidates and for in-group ones.

The organization has an arbitrary size. There are two horizontal groups, and two talent levels (we later generalize the talent distribution to a continuum). Organization's members enjoy linear quality and homophily benefits from their colleagues' attributes (we later generalize to non-linear and asymmetric homophily benefits). Quality benefits exceed homophily ones (the interesting case, as otherwise the majority only hires its in-group candidates). Section 2 fully characterizes pure-strategy Markov Perfect Equilibria. We show that MPEs satisfy the following properties: (a) equilibrium strategies are "canonical", meaning that hires are meritocratic except perhaps for tight majorities (when a minority appointment may lead to a loss of control); (b) a group is more inclined to be meritocratic if the other group also is (strategic complementarity); (c) in the symmetric case, the organization is either meritocratic or entrenched, and the two regimes coexist over a non-empty range of the quality-over-homophily-benefit ratio.

Section 2 then computes the welfares of current minority and majority members. The two equilibria, when coexisting, are Pareto-ranked with meritocracy dominating entrenchment, which enables us to make a selection and perform comparative statics and policy evaluation. A second measure of welfare is aggregate ergodic welfare. With this criterion as well, the meritocratic equilibrium, while delivering lower homophily benefits on average, dominates the entrenchment one.

Section 3 derives the implications of the three familiar public policies previously described. Section 4 relaxes the assumptions of the baseline model. It extends the analysis to a continuous quality space (4.1), homogamic evaluation capability (4.2.1), uncertain voting participation (4.2.2), and "anterooms for appointments", which can be external when rejected candidates may reapply or internal when junior members may be promoted to a senior position[2] (4.3). The paper concludes by discussing the related literature (Section 5) and avenues for future theoretical and empirical research (Section 6). Omitted proofs can be found in the Online Appendix.

2 Baseline 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 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 still being 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 talent or quality and takes one of two possible values, 0 (mediocre) or $\tilde{s}$ (talented), 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 horizontal type $X \in \{A, B\}$ exerts per-period externality $\tilde{b}_X > 0$ on members of the same type,[3] but not on members of the opposite type, and this regardless of their talent.[4]

We thus assume that each member derives utility from: (i) their colleagues' talent, i.e. the vertical attributes of members of the organization, and (ii) homophily over tastes: ceteris paribus, each member prefers colleagues who share their horizontal type. So, member ι of type $X \in \{A, B\}$ receives date-t flow payoff

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

[2]In hierarchical organizations, the oft-made observation that minorities experience difficulties in rising above a certain level suggests that meritocracy is more often violated at higher than at lower levels. Even if in-group favoritism contributes to discrimination against minorities, it is not a priori obvious that it should imply a lower rate of promotion for the latter (a "glass ceiling"). Nonetheless, Section 4.3.2 shows in the natural extension to a two-level organization that a glass ceiling results from control being located at the senior level

[3] The case $\tilde{b}_X < 0$, corresponding to *negative homophily* – e.g., envy towards the likes, preference for diversity or for a smaller in-group, etc. (see for instance Bagues and Esteve-Volart 2010) – can be accommodated in our model. See Online Appendix Q.

[4]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. Alternatively, 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.

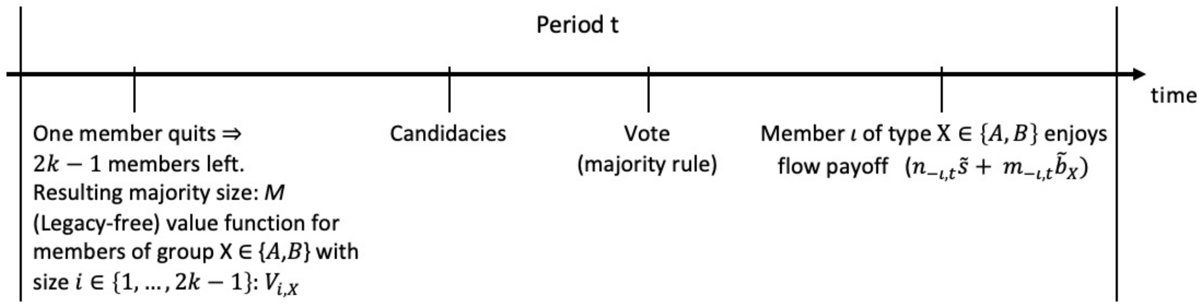

Figure 1: Timing.

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.[5]

In each period, there is at least one candidate of each type. Assuming then that there is exactly one of each type involves *no loss of generality* as all members of the organization always prefer the most talented candidate of a given horizontal type to any candidate of the same type but with lesser talent, and are indifferent if there are multiple "most-talented" candidates of a given horizontal type. Moreover, we assume that candidates apply to become members only once and that the candidates' types are observable prior to the vote – we will later relax these two assumptions.

Let $s \equiv \tilde{s}/[1-\delta_0(1-2/N)]$ denote the expected incremental lifetime contribution of a new talented (relative to mediocre) appointee to each incumbent member of the organization.[6] We similarly denote by $b_X \equiv \tilde{b}_X/[1-\delta_0(1-2/N)]$ the expected lifetime homophily benefit for an incumbent member of type $X \in \{A, B\}$ generated by a new in-group member.

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. Let $M \in \{k, k+1, ..., 2k-1\}$ denote the size of (number of individuals in) the majority. We say that the majority is *tight* if $M = k$.[7]

To make things interesting, we assume $s > b_X$ for all $X \in \{A, B\}$. Otherwise, if $b_X \geq s$, systematically voting for the majority candidate would yield the highest possible continuation payoff for majority X, and (in the absence of coordination failure) such a majority would always move towards perfect homogeneity.

Candidates' talents are i.i.d. across periods. We let $x \in (0, 1/2]$ 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 $\alpha \in [0, 1]$ denote the

[5] Our insights are unchanged if a talented member derives a "quality payoff" from her own talent.

[6] 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.

[7] We refer to a majority member as "he", to a minority member as "she", to a generic organization member as "they", and to the principal – whenever there is one – as "it".

probability that both are of talent s conditional on both being equally talented. The probability of an in-group (or out-group) candidate being of quality s is thus equal to $\overline{x} \equiv x+(1-2x)\alpha$.

Our equilibrium concept is pure-strategy Markov Perfection. Given the other group's strategy, all members of a given group $X \in \{A, B\}$ at date t have the same objective function. Moreover, if they have not exited by date $t + \tau > t$, they will have the same date-$(t + \tau)$ continuation payoff function as the other members of the same group, regardless of their respective cohort and talent. So, only the size M of the majority and its identity X are payoff-relevant in the sense of Maskin-Tirole (2001). Markov strategies therefore depend neither on time t, nor on the cohorts or talents of incumbent members.

In addition, we assume that each member votes as if they were pivotal, i.e., as if they alone chose the candidate. Hence, we ignore coordination failures in which, say, a majority member votes for an unfavored candidate because other majority members also do.[8] Consequently, at any date, all members of a given group vote unanimously to maximize their current-plus-continuation payoff.[9]

2.1 Equilibrium characterization and existence results

2.1.1 Majority's best response and strategic complementarities

The state variable for a majority of horizontal type $X \in \{A, B\}$ is its size $M \in \{k, ..., N-1\}$. Let us then study the "best response" for a majority of type $X \in \{A, B\}$ and size $M \in \{k, ..., N-1\}$ to the other group's strategy, summarized by the current majority's continuation value upon losing control (reaching size $k - 1$).

Since the present discounted value of benefits accruing from other incumbent members plays no role in an MPE, we do not include the legacy terms in the expression of the value functions. For any group size $i \in \{1, ..., N-1\}$ just before candidacies are declared (see Figure 1), we denote by $V_{i,X}$ the value function of an individual in group X: $V_{i,X}$ is the expected discounted value of flow payoffs brought about by colleagues who *will* be coopted later, in the current period and in future periods.

A majority member's continuation value at majority size $M \geq k$ is given by

$$b_X + s_{\text{maj}} + \delta\left[\frac{M}{N-1}V_{M,X} + \left(1 - \frac{M}{N-1}\right)V_{M+1,X}\right]$$

[8] The assumption that agents vote as if they were pivotal could stem in particular from a trembling-hand requirement as in Acemoglu et al. (2009), or from a coalition-proofness requirement among current members of the same horizontal group (majority/minority).

[9] Since we thus rule out coordination failures within the majority, the minority's behaviour is for now irrelevant (uncertain voting participation or identification of group allegiance will be considered in Section 4.2.2).

if the majority candidate with talent (expected lifetime contribution) $s_{\text{maj}} \in \{0, s\}$ is recruited in the current period, and by

$$s_{\min} + \delta\Big[\frac{M-1}{N-1}V_{M-1,X} + \Big(1 - \frac{M-1}{N-1}\Big)V_{M,X}\Big]$$

if the minority candidate with talent (expected lifetime contribution) $s_{\min} \in \{0, s\}$ is. The value function $V_{M,X}$ of a majority member at majority size M is the expectation of its continuation value over all current-period possible events (candidates' profiles and recruitment decisions).[10] The majority's choice between the two candidates is thus determined by the following comparison:

$$b_X + s_{\text{maj}} - s_{\min} + \delta\Big[\frac{M-1}{N-1}(V_{M,X} - V_{M-1,X}) + \Big(1 - \frac{M}{N-1}\Big)(V_{M+1,X} - V_{M,X})\Big] \lesseqgtr 0. \quad (1)$$

Canonical strategies. We will show that all pure-strategy, Markov Perfect best-responses are canonical, i.e. strategies such that:

(i) Members of the majority (all) vote for the majority candidate if the latter is at least as talented as the minority candidate,

(ii) When the minority candidate is more talented, members of a type-X majority, with $X \in \{A, B\}$, (all) vote for the majority candidate with probabilities $\{\sigma_X(M)\}_{M\in\{k,\ldots,N-1\}}$ with $\sigma_X(k) \in \{0, 1\}$ and $\sigma_X(M) = 0$ if $M > k$.

Intuitively, the assumptions of the basic model ensure that control can be retained simply by coopting a majority candidate when the majority is tight ($M = k$). In a canonical equilibrium, hires are meritocratic except perhaps for tight majorities. We will say that a type-X majority is

- *meritocratic* if $\sigma_X(M) = 0$ for all $M \geq k$;
- *entrenched* if it favors a mediocre majority candidate over a talented minority one only when the majority is tight ($M = k$), i.e. if $\sigma_X(k) = 1$ and $\sigma_X(M) = 0$ for all $M \geq k+1$.

For future reference, we will also say that a type-X majority is *entrenched at level l* if $\sigma_X(M) = 1$ for $M \in \{k, ..., k+l\}$, and $\sigma_X(M) = 0$ for $M \geq k+l+1$. Correspondingly, a type-X majority is *super-entrenched* (resp. *fully entrenched*) if it is entrenched at some level $l \geq 1$ (resp. $l = k-1$). Online Appendix A proves the following intuitive property:

Lemma 1. ***(Majority's best response and canonical strategies)*** *Fix $V_{k-1,X}$ in the feasible range ($\big[0, ((\overline{x}+x)s+(1-x)b_X)/(1-\delta)\big]$). The majority's best response to $V_{k-1,X}$ among pure Markov Perfect strategies is either canonical meritocracy or canonical entrenchment.*

[10] Moreover, any continuation value $V_{i,X}$ with $i \in \{1, ..., N-1\}$ thus lies in the interval $\big[0, ((\overline{x}+x)s + (1-x)b_X)/(1-\delta)\big]$.

Put differently, unless control is immediately at stake ($M = k$), the majority's best response is always to recruit the most talented candidate, breaking ties in favor of the in-group candidate.

Our next economic insight – the strategic complementarity of canonical strategies – builds on this observation. Formally, let us, abusing notation, denote by $V_{i,X}^{\mathrm{r,r'}}$ group X's continuation value function when it has size i and follows strategy r when it has the majority, with r = m if the latter is canonical meritocracy and r = e if it is canonical entrenchment, and the other group follows strategy r' $\in \{\mathrm{m}, \mathrm{e}\}$ when it has the majority. When $r = e$, control never switches and the other group's strategy is irrelevant, and so $V_{k,X}^{e,m} = V_{k,X}^{e,e}$.

When $r = m$ and the other group is entrenched, the "flow" payoffs in $V_{i,X}^{m,e}$ for $i \leq k-1$ are bounded above by $(\overline{x}+x)s + xb$ and are strictly lower than the bound when $i = k-1$, while the "flow" payoffs in $V_{i,X}^{m,m}$ for $i \lesseqgtr k-1$ are bounded below by $(\overline{x}+x)s + xb$ (and are strictly higher than the bound when $i \geq k$ if $x < 1/2$). This implies that $V_{k,X}^{m,m} > V_{k,X}^{m,e}$. And therefore,

$$V_{k,W}^{m,m} - V_{k,X}^{e,m} \;>\; V_{k,X}^{m,e} - V_{k,X}^{e,e}.$$

Consequently, group X is more inclined to be meritocratic when it has a tight majority if the other group is also meritocratic.

Proposition 1. ***(The strategic complementarity of canonical strategies)*** *A given group $X \in \{A, B\}$ is more inclined to be meritocratic (resp. entrenched) if the other group is itself meritocratic (resp. entrenched).*

2.1.2 The symmetric case

Except in the asymmetric extension of Section 2.3 and for expositional conciseness, we henceforth restrict our attention to the symmetric case in which both horizontal groups have the same homophily preferences: $b_A = b_B \equiv b$.

In the symmetric case, Lemma 1 and Proposition 1 together imply that generically, any pure-strategy MPE is canonical and symmetric,[11] i.e. that the behaviors of A and B majorities are the same, either canonically meritocratic or canonically entrenched. We therefore drop the subscript X in the value function. We simply call such equilibria *canonical* and refer equivalently to a meritocratic/entrenched majority or organization.

Before studying the existence regions of such equilibria, let us describe some properties of the value functions of majority and minority members under such strategies. Figure 2

[11] Suppose that in equilibrium, group X is meritocratic and group Y is entrenched. Necessarily, $V_{k,X}^{m,e} \geq V_{k,X}^{e,e}$ and $V_{k,Y}^{e,m} \geq V_{k,Y}^{m,m}$. However, for any $r, r' \in \{e, m\}$, $V_{k,X}^{r,r'} = V_{k,Y}^{r,r'}$ as payoffs are symmetric. The equality $V_{k,X}^{e,e} = V_{k,Y}^{e,m}$ would imply that $V_{k,X}^{m,e} \geq V_{k,X}^{m,m}$, which we know is impossible from the proof of Proposition 1.

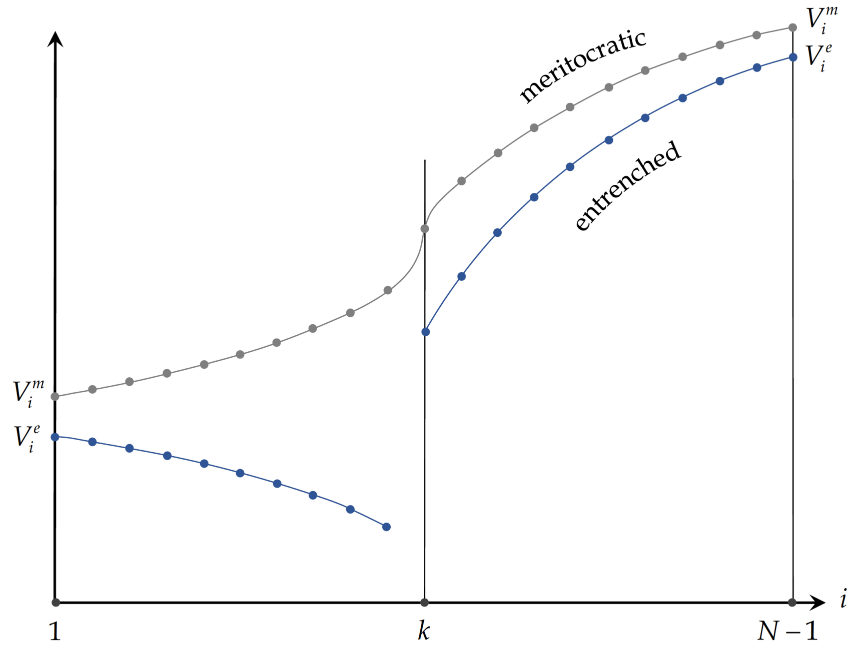

Figure 2: Properties of value functions in the meritocratic and in the entrenched equilibria.

illustrates the following lemma.

Lemma 2. ***(Properties of value functions in the meritocratic (m) and in the entrenched (e) equilibria)***

(i) (Majority value function) For $i \in \{k, ..., 2k-1\}$, V_i^e is strictly increasing in i and has strictly decreasing differences.[12] *Similarly, V_i^m is increasing in i and has decreasing differences, strictly so if and only if $x < 1/2$.*

(ii) (Minority value function) For $i \in \{1, ..., k-1\}$, V_i^e is strictly decreasing in i and has strictly increasing differences in i. 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 $x < 1/2$.

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

Intuitively, the three parts of Lemma 2 stem from the following observations. Firstly, in a canonical equilibrium, the majority always picks its "myopically favorite" candidate except in the entrenchment equilibrium when $M = k$, where "myopically favorite" refers to the choice the majority would make in the absence of future elections or, equivalently, if future hiring decisions did not hinge on the current one. The higher M is, the more remote the appointment of a

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

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

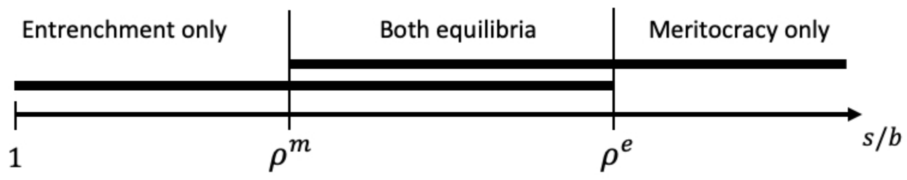

Figure 3: Existence regions for meritocratic and entrenched equilibria over the s/b line.

myopically suboptimal candidate (entrenchment equilibrium) and the more remote a possible loss of control (meritocratic equilibrium).

Secondly, for minority members, the impact of moving further away from the tight-majority state ($M = k$) depends on the equilibrium: in the entrenched equilibrium, the further away from minority size $k - 1$, the smaller the additional *loss* of getting one step closer to the majority's entrenched recruitment at $k-1$, whereas in the meritocratic equilibrium, the further away from minority size $k - 1$, the smaller the additional *benefit* of getting one step closer to possibly seizing control of the organization.

Thirdly, homophily induces a *benefit from control* for the majority whenever candidates have the same talent – as the majority can then pick its in-group candidate at no cost in terms of quality.[13]

Proposition 2. ***(The symmetric case: Canonical Equilibria)***

(i) All pure-strategy Markov Perfect equilibria are canonical and symmetric.

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

(ii) The entrenchment 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_0, $\partial\rho^m/\partial\delta_0 \geq 0$, and $\partial\rho^e/\partial\delta_0 \geq 0$.[14]

Figure 3 describes the existence regions over the line s/b for given x, δ. For s/b close to 1, selecting an untalented peer over a better qualified minority candidate comes at little cost for the majority, and there is a benefit from keeping control, so the majority is entrenched. As the ratio quality/homophily benefits s/b increases, the cooptation game moves from a (bounded) region where only the entrenchment equilibrium exists, to an intermediate (bounded) interval where, due to strategic complementarities, both equilibria coexist.[15] As s/b continues to increase, it reaches the (half-line) region where only the meritocratic equilibrium exists.

[13] The benefit from control persists with a continuum of vertical types (see Section 4.1) as the majority then reaps a homophily benefit when recruiting its in-group candidate against a slightly more talented out-group candidate.

[14] Furthermore, for δ_0 small, ρ^{m} and ρ^{e} increase with the size of the organization $N = 2k$.

[15] As usual, there is then a third, mixed-strategy equilibrium with $\sigma(k) \in (0, 1)$. (We focus on pure-strategy equilibria throughout the paper. Besides, as will be clear shortly, this mixed-strategy equilibrium is dominated by the pure-strategy meritocratic equilibrium.)

As the discount factor increases, the existence region of the meritocratic equilibrium shrinks while that of the entrenchment equilibrium widens. This comparative static is 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. By contrast, $\rho^{\mathrm{m}} > 1$ whenever $x < 1/2$.

2.2 Welfare

2.2.1 Non-ergodic member welfare

We first consider current members' welfare, defined as their expected discounted surplus (from current and future hires), at any given legacy and period, therefore computed from the continuation payoffs V_i. We refer to this welfare notion as "non-ergodic member welfare".

The next Proposition shows that, when they coexist, the meritocratic equilibrium is preferred to the entrenched one by all members of the organization. Intuitively, at any given majority size, minority members 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.[16]

Proposition 3. ***(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 current members of the organization to the entrenchment equilibrium.*

As a consequence, from the perspective of current members, the meritocratic equilibrium Pareto-dominates the entrenchment equilibrium whenever they coexist.

2.2.2 Ergodic aggregate welfare

We now draw an aggregate-welfare comparison between entrenchment and meritocracy in their respective ergodic distribution from the perspective of a principal or third-party putting at least as much weight on quality as on homophily benefits. Denoting by S the organization's ergodic per-period aggregate quality, by B the ergodic per-period aggregate homophily benefits, the principal's objective writes

$$W \equiv qS + B$$

[16] Regardless to the regime $\mathrm{r} \in \{\mathrm{m}, \mathrm{e}\}$, the majority faces an optimal stochastic control problem with boundary value V_{k-1}. All valuations V_{k+l}, with $l \geq 0$, are therefore non-decreasing functions of V_{k-1}.

where $q \geq 1$ is the (relative) weight put by the principal on quality relative to homophily. The "no-externality-on-third-parties case" $q = 1$ corresponds to the maximization of (ergodic) total member surplus. But it often makes sense to assume that $q > 1$: homophily benefits are fully appropriated by the members, while talent yields benefits for both members and their organization or society (taxes, innovation, prestige, etc.).

We first describe the ergodic distributions of majority sizes. Since, by convention, 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. Let us index the end-of-period majority size by $i \in \{k, ..., N\}$. Let ν_i^{r} denote the ergodic probability of majority size i at the end of a period in regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$ (see Online Appendix E for their expressions). The next Lemma shows that entrenchment leads to larger majorities, as intuitive:

Lemma 3. ***(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,[17] aggregate per-period expected quality S^{r}, $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$ is

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

Unsurprisingly, the ergodic quality of a meritocratic organization 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 expected homophily benefit B^{r} writes

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

[17] 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,

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

where $\overline{x} = x + (1-2x)\alpha$ is the probability of an in-group (or out-group) candidate being of type s. Similarly for the entrenchment equilibrium,

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

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 3.

The following result compares the two *laissez-faire* equilibria's ergodic welfares. It intuitively stems from $s \geq b$ and quality benefits accruing to all members while homophily ones profit only the in-groups.

Proposition 4. ***(Ergodic per-period aggregate welfare)*** *For any $s \geq b$, $W^m > W^e$, i.e., the meritocratic selection rule dominates the entrenchment one in terms of ergodic per-period aggregate welfare.*

The collegial bias against meritocracy. Let us emphasize an important asymmetry between entrenchment and meritocracy. Indeed, Propositions 2 and 4 imply that while there exists an entrenched equilibrium whenever entrenchment is desirable in terms of ergodic aggregate welfare, there exists a non-empty range of values of s/b such that meritocracy is desirable and yet equilibrium behavior leads to entrenchment.[18] We interpret this discrepancy as a *collegial bias against meritocracy* when talent is scarce and political strength lies in numbers.

Measuring meritocracy. We defined *meritocracy* as the recruitment of the most talented. Considering an organization's ergodic state, one may thus measure its degree of meritocracy by computing the probability that the (or "a" in case of a tie) most talented candidate is recruited. In the canonical-meritocracy and canonical-entrenchment equilibria, this probability is equal respectively to 1 and to $1 - \nu_{k+1}^{\mathrm{e}} \frac{k+1}{N} x$ (while with full entrenchment, which prevails in equilibrium whenever $s < b$, it would be equal to $1 - x$).

2.3 General homophily benefits

Let us first consider two extensions regarding the shape of homophily benefits, relaxing symmetry and linearity.

Asymmetric homophily benefits: It takes two to build meritocracy. Returning to linear homophily benefits, suppose that type-A agents have stronger homophily preferences

[18] This result extends to the case $s/b < 1$: $W^{\mathrm{m}} > W^{\mathrm{e}}$ whenever $s/b > \rho^W$ for some $\rho^W < 1$. Indeed, for $s/b < 1$, full entrenchment is the unique equilibrium even when it is dominated in terms of ergodic aggregate welfare by canonical meritocracy and/or canonical entrenchment (with meritocratic hires whenever $M \geq k+1$). In other words, regardless of whether s/b is higher or lower than 1, there exists a range of values of s/b such that a *certain degree of meritocracy* is desirable yet fails to exist in equilibrium.

than type-B agents, i.e. that $b_A > b_B$. So, from the point of view of meritocracy, group A is the "weak link".

Proposition 2'. ***(Asymmetric homophily benefits)*** *With asymmetric homophily benefits b_A, b_B such that $b_B < b_A < s$,*

(i) The canonical meritocratic equilibrium exists if and only if $s/b_A \geq \rho^m$.

(ii) The canonical entrenchment equilibrium exists if and only if $s/b_B \leq \rho^e$,

(iii) If homophily benefits are sufficiently dissimilar (so that $\rho^e b_B < \rho^m b_A$), the unique MPE in pure strategies when $\rho^e b_B < s < \rho^m b_A$ is such that type-A members follow the canonical entrenchment strategy, while type-B members follow the canonical meritocratic strategy. On path, type-A members eventually have the majority and the equilibrium becomes observationally equivalent to canonical entrenchment.

(iv) If they are sufficiently similar (so that $\rho^m b_A < \rho^e b_B$), the meritocratic and entrenchment equilibria coexist over a non-empty range of qualities s. Whenever they do, the meritocratic equilibrium Pareto-dominates the entrenchment one.

As Proposition 2' shows, it takes two to build meritocracy, and it takes only one to destroy it. Indeed, when facing a rival group with strong homophily preferences, an otherwise meritocratic group either anticipates the entrenched behavior of its rival, and thus deviates from meritocracy to "preemptive entrenchment", or sticks to meritocratic recruitments, only to eventually lose control to the other group who then entrenches its majority.[19]

Non-linear homophily benefits. Convex homophily benefits arise for instance when facilities or regulations must be added to accommodate the existence of a minority, or when a group's reaching a critical size delivers additional opportunities to its members, e.g., because of supermajority clauses for some decisions.[20] Conversely, concave homophily benefits arise if there are decreasing returns to having one more in-group member (e.g., limited time for "horizontal" interactions) or increasing returns to having one more out-group member (e.g., benefits from diversity).

Let us from now on return to the symmetric case ($b_A = b_B = b$). 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 depends

[19]Put differently, an increase in b_A can have long-term consequences on the organization's dynamics, whereas an increase in b_B (still below b_A) has at most short-term consequences.

[20]The case in which each period, a non-hiring decision is subject to a supermajority rule is indeed 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).

on future hirings under a non-linear homophily benefit, the meritocracy-vs-control trade-off remains. Let $\tilde{\mathcal{B}}(i)$ denote the flow homophily benefit enjoyed by a member whose in-group has size i (thus, in the linear case, $\tilde{\mathcal{B}}(i) \equiv (i-1)\tilde{b}$).

Proposition 2". ***(Non-linear homophily benefits)***

(i) With strictly concave homophily benefits $\tilde{\mathcal{B}}$, symmetric MPEs are still either meritocratic or entrenched if $\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k) \leq \tilde{s}$, and are super-entrenched if $\tilde{\mathcal{B}}(k+2) - \tilde{\mathcal{B}}(k+1)$ is sufficiently large.

(ii) With strictly convex homophily benefits, when there exists a threshold size such that $\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M) < \tilde{s}$ for any M below the threshold and $\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M) > \tilde{s}$ for any M above the threshold,[21] then in any symmetric MPE, there exists a threshold size such that recruitments are entrenched for majority sizes above the threshold. On path, equilibria eventually become observationally equivalent to full entrenchment.

For expositional simplicity, we henceforth resume the case of linear, symmetric homophily benefits and we restrict our attention to symmetric equilibria. (As we have shown, all equilibria are symmetric in the basic model.)

3 Policy

We next investigate the consequences of different interventions. To perform the policy analysis, we need to select an equilibrium in the multiple-equilibria region (our insights however do not depend on this particular selection). Motivating our choice is our previous result that whenever meritocracy and entrenchment coexist, meritocracy Pareto-dominates entrenchment.

Assumption. ***(Equilibrium selection).*** *Whenever two equilibria coexist, coordination occurs on the meritocratic one. So, under laissez-faire, entrenchment prevails if and only if $1 < \rho < \rho^m$.*

Public interventions depend on the nature of the principal's information. The principal may use knowledge about vertical attributes either to *overrule majority decisions* or to *provide quality-based incentives*: A provost uses external letters or a visiting committee to assess the quality of a department or candidates' talent; a government designs a research assessment exercise to evaluate a university or its components. Alternatively, the principal may observe horizontal types: *Affirmative action policies* are based on gender, race, disability, or religion, but not necessarily on a measure of talent.

[21] The threshold can be equal to $N-1$ for δ sufficiently low, and must be weakly below $N-2$ otherwise.

Next, we need to specify the principal's objective function. As for our earlier computation of ergodic welfare, we will base welfare on the surplus that the organization generates: $W = qS + B$, where $q \geq 1$ to accommodate the possible presence of externalities of quality on the broader society. In the case of incentives, we may subtract their total cost T, and so, $W = qS + B - \xi T$, where the magnitude of the shadow cost of public funds $\xi \geq 0$ depends on whether the average compensation can be decreased accordingly when T increases.

A comment regarding our objective function is in order. Besides efficiency, other familiar policy motivations include the quest for justice or the benefits from minority role models. In our model, such goals are naturally included in q: for, policies that encourage meritocracy in the sense of selecting the most talented also operate toward benefiting minorities.[22]

Previewing the formal analysis, the consequences of the interventions we investigate have three drivers:

(1) A reduction in the value of decision rights: majorities cannot optimize as efficiently when they face external constraints. This lower value from control reduces the appeal of entrenchment (*loss-of-control-value effect*).

(2) The fear of an involuntary loss of control due to discretionary interventions in cooptation decisions may encourage the current majority to build a buffer against such majority transfers, i.e., to super- or fully entrench itself (*precautionary-buffer effect*).

(3) When the organization members are rewarded for overall quality, vertical considerations are strenghtened with respect to horizontal ones, and thus quality-based rewards favor meritocracy (*higher-quality-relevance effect*).

Remark: Internal equality favors external meritocracy. An interesting illustration of the loss-of-control-value effect crops up when an organization's majority controls not only recruitments, but also the sharing of spoils among members (allocation of the organization's surplus). Mandating equal treatment of the organization's members reduces the value of such decision rights, lowers the majority's payoff and raises the minority's one for a given recruitment policy, and thus fosters meritocracy. We thus infer that non-discrimination requirements in non-appointment decisions favor meritocracy.

3.1 Quality-based interventions

Assuming that the principal has only talent information, we consider two policies for the principal: (i) stepping in to choose the new member (which does not assume commitment by

[22]This need not hold in general. Suppose, say, that for legacy/discrimination reasons a minority group has few highly qualified candidates. Then, promoting justice and generating role models can in the short run work against making the organization more efficient. Such considerations could be embodied into the model. For expositional simplicity, we focus on the current case in which incentivizing the hiring of the most talented also benefits the minority.

the principal), and (ii) rewarding quality (which does). These two policies feature respectively the loss-of-control-value effect and the precautionary-buffer effect for the former, and the higher-quality-relevance effect for the latter.

Discretionary overruling of majority decisions. Suppose that in each period, the majority selects among the two candidates, yet the principal can then overrule the majority and pick the losing candidate. None of the players (principal, majority, and minority) can commit. In each period, the principal learns the quality of the current candidates (or at least their quality differential) with probability λ, and remains uninformed with probability $(1-\lambda)$.

We will look for equilibria (a) with level of entrenchment $l \geq 0$, and (b) in which the principal overrules the majority if and only if it is informed that the majority is violating meritocracy.

The principal's willingness to intervene is tied to the availability of evaluative information. In particular, in the absence of commitment, it is an equilibrium for the principal not to intervene without (talent) information.[23] Hence, for $\lambda = 0$, the meritocratic and entrenchment equilibria exist for the same parameter values as in the absence of intervention. At the other extreme, when $\lambda = 1$, the principal can (and will) select the best candidate in each period, and there is no real "cooptation". Hence, let us assume that $0 < \lambda < 1$. Regardless of λ, the existence condition 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. By occasionally overruling the majority, it can impose the meritocratic choice. But the majority may become wary of losing control when $M = k$ and may thus 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 the ability to overrule the majority systematically backfires by generating full entrenchment for s/b close to 1.[24]

Proposition 5. ***(Unintended effects of discretionary quality-based interventions)*** *Suppose $x < 1/2$ (there are benefits from control).*

(i) Fix $\lambda \in (0, 1)$. The possibility of an informed overruling of majority decisions (with a

[23] See Online Appendix I. Intuitively, (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.

[24] A company's board of directors may be subject to such interventions by the company's shareholders. Our results thus show that the shareholders' threat of stepping in and appoint a minority candidate may backfire and induce the current majority to further entrench itself. All other things equal, a board facing milder threats would be more prone to opt for meritocratic hiring. In practice nonetheless, shareholders' activism may have an ambiguous impact as they may also influence – via their trading and voting decisions – the directors' *rewards for quality*, which is the intervention we study next.

strictly positive probability) results in full entrenchment for any $s/b \in [1, \rho]$ *for some* $\rho \in (1, \rho^m)$.

(ii) *Fix* s/b *close to* 1*. Suppose the principal values quality enough (*q *large). For* λ *in an intermediate range, the principal achieves a higher ergodic quality if it can commit not to intervene.*

Rewarding quality. We now assume that the principal implements a quality assessment exercise after each period's election with probability corresponding to a Poisson process of rate η. 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 without loss of generality that the bonus earned at date t is immediately paid to the organization.[25] For the sake of simplicity, we also assume that 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.[26]

Letting T denote the ergodic expected per-period transfer and ξ its shadow cost, the ergodic welfare function in the presence of transfers becomes: $W = qS + B - \xi T$.

Proposition 6. ***(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* W *if and only if* $s/b \in [\rho_\xi, \rho^m)$.

The intuition behind Proposition 6 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 appetence 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.[27] Thus, a rule that applies to all organizations is likely to lack efficiency.

[25] Alternatively, we could have assumed that the bonus is split across several periods. Yet, frontloading the bonus is more efficient. Indeed, because members may quit, and thus $\delta \leq (N-1)/N < 1$, frontloading the bonus maximizes the incentive for good recruitment.

[26] 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, for η, y sufficiently high, the ratio s^+/b is sufficiently high for the organization to reach the region where the canonical meritocratic equilibrium exists.

[27] The optimal transfer is equal to 0 for s/b below a certain threshold (which increases with the cost of public

3.2 Affirmative action

Suppose that the principal mandates diversity by setting a "representation threshold" – i.e. imposing that the minority count at least R members at the end of any given period. Since it is suboptimal for the principal to impose parity,[28] 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 lead to a "constrained meritocratic" equilibrium (in which the recruitment choice is meritocratic except perhaps when $M = N - R$ at the moment of the vote), which might actually benefit the principal.[29] The policy however is costly when the meritocratic decision would select the majority candidate, but the minority quota is binding, forcing the recruitment of the minority candidate.

Proposition 7. ***(Affirmative action: Representation thresholds)***
(i) Existence region. *Affirmative action in the form of a representation threshold $R \leq k-1$ expands the region for which majority alternance prevails in equilibrium.*
(ii) Ergodic aggregate welfare. *When $s/b \geq \rho^m$, affirmative action comes at a cost, both in terms of efficiency and homophily, and reduces welfare. When $s/b < \rho^m$ and affirmative action induces an otherwise entrenched organization to become meritocratic, there exists a cut-off in the correlation of candidates' vertical types such that affirmative action dominates laissez-faire if and only if the correlation is below the cutoff, and is dominated otherwise. The more ambitious the affirmative action, the lower the cut-off.*[30]

funds ξ), jumps discontinuously strictly above zero at this threshold, and then decreases with s/b above the threshold, down to zero when $s/b = \rho^{\mathrm{m}}$.

[28] Suppose that the principal imposes parity (so at the end of any 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.

[29] When s/b is very high, the efficiency loss at $M = N - R$ becomes extremely costly and majority members may 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. *Meritocracy with reverse favoritism* may thus arise in equilibrium: majority members vote for their candidate if and only if he is strictly more talented than the minority candidate. How relevant is such reverse favoritism? For s/b high, meritocracy is likely to prevail in the organization and regulators unlikely to intervene on an *ad hoc* basis. But an economy-wide affirmative action rule would apply even to organizations that would otherwise be meritocratic, giving rise to reverse favoritism.

[30] Namely, we show that: (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 $x_{\mathrm{AA}}(R) \in (0, 1/2)$ such that for any $x \in (0, x_{\mathrm{AA}}(R))$ (resp. $x \in (x_{\mathrm{AA}}(R), 1/2)$), 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 $x = x_{\mathrm{AA}}(R)$). The cutoff $x_{\mathrm{AA}}(R)$ 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.

Interestingly, by reducing the value for majority members of controlling recruitments, affirmative action can switch the organization's regime from entrenchment to meritocracy, thereby having an impact on recruitments above and beyond the immediate hiring constraint when the threshold is reached. But the organization must trade off the benefit and the cost of affirmative action. This may be difficult, in particular as the policymaker may be imperfectly informed of the organization's natural propensity to be meritocratic (and this difficulty may be aggravated by the organization's strategic response to the regulation if the intervention is organization-specific).

4 The collegial bias against meritocracy: Further drivers

This section considers robustness results and extensions, causing organizations and their members to depart from the (meritocratic and entrenched) canons we described in Section 2. More pervasive and/or more intense forms of entrenchment arise, indicating that in practice, one can expect (significantly) fewer meritocratic recruitments than described by these canons.

4.1 A continuum of vertical types

We have assumed so far that talent can take only two values. When talent is smoothly distributed in $\mathbb{R}_+$, for the natural generalization of canonical equilibria developed below, full meritocracy never prevails, as the majority always prefers an in-group candidate over a slightly more talented out-group candidate. But, as we will see, we can still order equilibria in terms of their "level of meritocracy". Our previous insights generalize: (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.

Generalizing canonical equilibria to arbitrary talent distributions, equilibria can be described as a sequence of strictly positive cut-offs $(\Delta_M)_{M\in\{k,...,N-1\}}$ such that 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 $\hat{s}-s > \Delta_M$. We show in Online Appendix L that in any such equilibrium, $\Delta_M > b$ for any $M \in \{k, ..., N-1\}$. Intuitively, in-group recruiting when $b > \hat{s} - s$ yields a double dividend – a larger homophily payoff and a tighter grip on the organization –, and thus for a minority candidate to be considered by the majority, her talent must exceed the majority candidate's by strictly more than the homophily benefit: $\hat{s} - s > b$.

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.

Definition. *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 in* $[0,+\infty)^2$ *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* $\overline{s} = \sup(\hat{s}-s) \in (b, +\infty]$.

The set $\mathcal{G}$ includes the set of (full support) continuous joint symmetric distributions with finite-mean marginals. 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 8. ***(A continuum of vertical types)*** *Assume talents are distributed according to a joint distribution* $G \in \mathcal{G}$*. Any symmetric MPE described by a sequence of cut-offs* $(\Delta_M)_{M\in\{k,\ldots,N-1\}}$ *is such that* $\Delta_M > b$ *for any* M*, and that the sequence* $(\Delta_M)_M$ *is strictly decreasing: a stronger majority discriminates less than a weaker majority.*

Moreover, whenever they coexist, any two such equilibria 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 at any majority size by all current majority members, and for* δ *small, by all current minority members as well.*

4.2 Further drivers of super-entrenchment under laissez-faire

Returning to the binary-talent case, the most obvious case for super-entrenchment is $s \leq b$, which trivially leads to full entrenchment. Section 2.3 noted that, even for $s > b$, non-linear homophily benefits may lead to super-entrenchment. Departing from laissez-faire, Section 3 showed that some well-meaning interventions may have the unintended consequence of incentivizing the majority to be super-entrenched. Let us now describe two other drivers of super- and full-entrenchment under laissez-faire: homogamic evaluation capability and uncertain voting participation or identification of group allegiance.[31]

[31] Further away from our environment, the organization's horizontal homogeneity can stem from a lack of attractiveness to minority candidates. Indeed, relaxing our assumption of exogenous candidacies, individuals may apply only if they find the organization sufficiently attractive and their chances of being recruited sufficiently high. For conciseness, we focus in this paper on control concerns and refer to Moisson-Tirole (2024) for a study of attractiveness concerns and the "organizational spirals" they induce.

4.2.1 Homogamic evaluation capability

We assumed 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 conversely.

When only in-group evaluation is feasible, 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

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

Let us focus on the case of "pessimistic expectations" (or negative stereotypes): $s^{\dagger} \leq b$. 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, this case would also be more likely if the majority members had a negative stereotype about minority members' talent.] When $s^{\dagger} \leq b$, 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.

Online Appendix M studies the more complex case in which the minority candidate is in expectation preferred to an untalented majority candidates ($s^{\dagger} > b$). Then, analogues of the canonical (meritocratic and entrenched) equilibria exist, in which the minority candidate is given the benefit of the doubt except perhaps when the majority is tight ($M = k$).[32]

Proposition 9. ***(Canonical equilibria with homogamic evaluation capability)*** *If $s^{\dagger} \leq b$ (pessimistic expectations/negative stereotype), the majority coopts only candidates of the in-group and therefore becomes homogeneous. Homogamic evaluation capability then lowers the ergodic per-period aggregate welfare relative to perfect information.*

Remark: Cheap talk. One may wonder whether communication could help the majority select a candidate. The answer is that, in the complete absence of commitment, for $x^{\dagger} \leq 1/2$, one-shot 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,

[32] Online Appendix M further shows that the meritocratic and entrenchment equilibria with homogamic evaluation capability yield a lower ergodic aggregate welfare than their perfect-information counterparts.

and so any message sent to the majority is necessarily uninformative. Notwithstanding, more elaborate communication or voting mechanisms than majority rule, designed to incentivize the minority to reveal (some information about) the talent of its in-group candidate, may mitigate some drawbacks of homogamic evaluation capability.[33]

4.2.2 Uncertain voting participation or identification of group allegiance

We have assumed so far that all members of the organization vote and that they vote as expected. Relaxing the first assumption, we note that absenteeism, whether due to illness or alternative obligations, may incentivize groups to secure majorities of more than one vote so as to minimize the probability of a control switch. Consequently, with uncertain voting participation, even large majorities may find it optimal to stand in the way of talented minority candidates.

Returning to symmetric evaluation capability, 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 choice prevails.[34] 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, with the majority being certain to win for sufficiently large majority sizes (clearly so for $N-1$):[35]

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

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 (2) 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 (2) may not be warranted if voting participation is strategic rather than caused by exogenous events.

We look for monotonic (in the sense that a stronger majority makes more meritocratic

[33] Regarding such voting mechanisms, see for instance Jackson-Sonnenschein (2007), Casella et al (2008), Lipnowski-Ramos (2020). We leave a detailed study of the optimal voting mechanisms in our environment for future work.

[34] 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.

[35] 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.

recruitments), pure-strategy symmetric MPEs, which indeed exist. In contrast to the baseline model, the minority's strategy now matters at any majority size M at which $\Lambda(M) > 0$. When looking for level-l super-entrenchment equilibria, we now look for equilibria in which (a) the majority is entrenched to level l and (b) the minority always votes for its in-group candidate whenever it is pivotal with a strictly positive probability, i.e. whenever $M \leq k + l - 1$.

Proposition 10. ***(Absenteeism and super-entrenchment)*** *Let Λ satisfy (2) and $x < 1/2$. For s/b sufficiently close to 1, super-entrenchment at level l is the unique symmetric MPE such that a stronger majority makes (weakly) more meritocratic recruitments.*[36] *In particular, if $l = k - 1$ as in the Bernoulli case, the possibility of absenteeism triggers full-entrenchment for s/b sufficiently close to 1.*

When Λ satisfies (2) with $l < k-1$, the majority is "safe" at any majority size $M \geq k+l+1$ as it will still control the outcome with probability 1 in the next period. Therefore, meritocratic recruitments are optimal at these majority sizes.

Remark: Fighting absenteeism. Interestingly, under the conditions of Proposition 10, strong in-group discipline, inasmuch as it reduces voting uncertainty, makes hiring more meritocratic. More generally, any policy intervention curbing absenteeism may make all members better off by reducing the degree of entrenchment.

Imperfect identification of group allegiance. Our modelling of uncertain voting participation also applies to imperfect identification of group allegiance. As an illustration, let us introduce the possibility that a candidate be able to mascarade as belonging to the other group and thereby be elected. Namely, let us assume there is a probability $\vartheta \in (0, 1/2)$ that the best candidate of the majority group[37] is incorrectly identified (tagged as belonging to majority group, when actually belonging to the minority group). To avoid having to consider complicated coming-out 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 above argument on uncertain voting participation (replacing the probability of the majority losing the vote with the probability of recruiting a minority candidate incorrectly identified), there exists a non-empty neighbourhood of 1 such

[36] Furthermore, for s/b sufficiently close to 1, in any symmetric MPE, the majority is entrenched when it has size $k + l$.

[37] We further assume that all candidates of the majority group are equally "unreliable" (incorrectly identified with the same probability). Otherwise, an untalented yet fully "reliable" candidate (i.e. identified as perfectly belonging to the majority) might then be preferred to a talented yet "unreliable" candidate.

that for s/b in this neighbourhood, the only monotone equilibrium is the full-entrenchment 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 belonging to a given horizontal group still prefer in-group colleagues all else being equal, but that the intensity of candidates' homophily-vs-quality preferences varies and is not observable. So a majority recruit may for instance put a much higher weight on talent relative to homophily than the average majority member[38] and therefore resist the entrenched strategy. Anticipating this possibility, the majority may again want to be super-entrenched, so as to minimize the probability of a switch in control.

4.3 Anterooms for appointments

We have so far viewed the appointment process as an organizational choice between recruiting a candidate 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 junior members for possible promotion to senior positions. This section analyzes these two possibilities, which display several similarities.

4.3.1 Candidates can re-apply

We investigate the consequences of unselected candidates being able to re-apply. Unsuccessful candidates keep re-applying until they are recruited.[39] For the sake of exposition, we make a further simplifying assumption: $\alpha = 0$, so that in any period, the new majority and 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 immediately hired and so the ability to re-apply is irrelevant on an equilibrium path. However, the knowledge that talented minority candidates will reapply lowers the cost of entrenchment and thus favors discrimination.

Proposition 11. ***(Reapplying for membership)*** *Assume $\alpha = 0$. Entrenchment yields the majority a higher value function 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*

[38] For example, a small fraction of majority candidates might have homophily benefit zb, where $z < 1$, and thus favor the meritocratic strategy over the entrenched one, despite their colleagues in the majority favoring the latter over the former.

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

majority of turning down her application. Moreover, the existence region for the meritocratic equilibrium shrinks when the organization can store applications.

4.3.2 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 e.g., in Bertrand (2018), we here investigate whether the desire for the dominant group to retain control might be a factor.

Online Appendix P considers a (large) two-tier organization with (many) senior and junior positions. At each point in time, a fraction of seniors exogenously departs and is replaced by juniors promoted to seniority. A fraction of juniors exogenously quit the organization as well. Flows out of the junior pool are offset by new recruitments. Seniors have control over hiring and promotion decisions.

We say that a glass ceiling exists if the probability of promotion of talented majority member is higher than the one of a minority member. Even if majority dominance and favoritism contribute to hiring discrimination against minorities, it is not a priori obvious that they imply a lower rate of promotion for the latter within the organization. Indeed, hiring discrimination implies that minority recruits are fewer and more talented than majority ones. We nonetheless show that a glass ceiling arises in our framework[40] provided that at least one of the following two effects operates:

- *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).
- *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.

Proposition 12. ***(Glass ceiling)*** *In the hierarchical organization's steady state, hiring at the junior level is fully meritocratic.*[41] *By contrast, there exists a glass ceiling for minority juniors: A talented minority junior is less likely to be promoted than a talented majority junior.*

[40] This environment can be enriched in interesting ways. See Online Appendix P.

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

5 Related literature

Empirical evidence. There is growing evidence that meritocracy may not prevail even in organizations that are incentivized to behave efficiently. Zinovyeva-Bagues (2015) shows 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 evaluation bias dominates the informational gain. Bagues et al. (2017) find that in (Italian and Spanish) scientific committees, male evaluators become less favorable to women if a woman joins the evaluation committee, suggesting horizontal control concerns from male evaluators. 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), and that the overruling of test scores ranking lowers quality, suggesting either poor judgement or (more interestingly for us) homophily objectives. Relatedly, Moreira-Pérez (2022) study the consequences of the 1883 Pendleton Act, which mandated exams for some employees in the largest US customs-collection districts, and find that although the act improved targeted employees' professional background, it incentivized hiring in exam-exempted positions, distorting districts' hierarchical structures. This countervailing response echoes our study of the unintended consequences of well-meaning policy interventions. 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.[42]

Theoretical 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 may 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

[42] The gender gap and glass ceiling 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).

wants to boost minority workers' incentives to invest in skills. Affirmative action gives the minority prospects and, if modest, boosts its incentives, but if extensive, creates a "patronizing equilibrium" and reduces incentives. 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.

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[43]. 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 4.2.1 on homogamic evaluation capability bears resemblance with Board et al. (2020), which assumes that talented people are better at identifying new talents, hence deriving rich dynamics. Section 4.2.1 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.[44]

[43]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.

[44]Moldovanu-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

Trade-off between talent and like-mindedness. Cai et al. (2018) analyzes the dynamics of a three-member club. Players are characterized by a vertical and a horizontal type, but unlike in our paper, homophily benefits are constant-sum (they stand for the sharing of spoils), while they are not in our model. Sections 2 and 4.1 generalize the analysis of Cai et al. to an arbitrary-size organization, arbitrary homophily benefits and a larger set of talent distributions, deriving new insights. While Cai et al.'s model includes costly search for candidates,[45] our model allows for a much larger scope of inquiry. Notably, in contrast with Cai et al., we investigate super-entrenchment and explore its drivers – in particular, we show that super-entrenchment can stem from non-linear homophily benefits, uncertain voting participation or homogamic evaluation capability, or else be the unintended consequence of several policy interventions. Moreover, while Cai et al. focus on finding the optimal voting rule in a three-member club,[46] we study a distinct and wide set of familiar policy interventions, including affirmative action, quality-based rewards, discretionary overrulings of majority appointments, curbing absenteeism, etc. In particular, we describe how such policies generate two conflicting effects: the *loss-of-control-value* effect and the *precautionary-buffer* effect.

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

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-Polborn (2011), similar backgrounds (homophily dimension) facilitate the estimation of others' 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. Modelling a single organization searching for candidates, Fershtman-Pavan (2021) show that if the evaluation of minority candidates is noisier than the one of majority candidates, then "soft affirmative action policies" tilting the search technology in favor of minority candidates in the candidate pool can backfire and actually reduce the likelihood of a minority candidate being recruited.

[45] 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, in Cai et al., members engage in costly search for candidates and as current members are not infinitely lived and thus do not enjoy the benefits of quality recruitment as long as the organization, they underinvest in search. 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.

[46] Our model also allows for a general investigation of voting rules in clubs of arbitrary size. Consider for instance supermajority voting rules. Suppose that a (completely uninformed) principal mandates that, to be elected, a candidate must receive at least $k+l$ votes, where $l \geq 1$. If no candidate reaches the election threshold, the principal picks one among the two candidates at random. As intuitive from our analysis, such a super-majority voting rule jeopardizes the majority's control when it has a size below the threshold. Unsurprisingly, it can be shown that for $x < 1/2$, for s/b sufficiently close to 1, super-entrenchment at level l is the unique symmetric MPE such that a stronger majority makes (weakly) more meritocratic recruitments.

efficient mentoring. Control is not a focus of their paper, unlike ours.

6 Future research

This paper studies homophily-induced control concerns in collegial organizations. It provides rich and testable insights as to where and when such concerns lead to violations of meritocracy. It investigates several potential remedies, identifying conditions for their effectiveness and warning about their (possibly dramatic) unintended consequences.

On the positive front, this paper's insights belong to two main themes. Firstly, *meritocracy is at risk whenever control stakes are high.* For instance, distrust of the outgroup jeopardizes meritocracy as each group is more eager to cling to power if it suspects the other group would not fulfil its part of the meritocratic deal. Relatedly, an organization is only as meritocratic as its less-meritocracy-prone group (its weakest link). Similarly, longer tenures within the organization heighten control stakes and foster entrenchment. Secondly, *meritocracy is at risk whenever control itself is at stake*, i.e. either frail, uncertain or impeded. For instance, larger majorities are more meritocratic than thinner ones; majorities with stronger group discipline are more meritocratic than majorities unable to prevent absenteeism or turncoats; independent majorities are more meritocratic than majorities exposed to outside overrulings.

On the normative front, how can discrimination be fought? Our first prediction is that dominant groups will play cat-and-mouse with the social planner, possibly making well-meaning policies backfire not only for society, but also for minorities. If control is at stake, the dominant group can stuff the organization with its candidates much more than it would have done in the absence of public intervention. Consequently, there may be no middle way and the principal may have to run the organization if it starts meddling with appointments at all. Our second prediction is that lowering control stakes can be effective. For instance, limiting the majority's power (while leaving control unchallenged) by mandating equal treatment of members, or by setting a minority representation quota fosters meritocratic behavior. So does raising non-control-related stakes via targeted subsidies. As these interventions entail different costs – cost of public funds in case of transfers, cost of ensuring that equal treatment of wage and working conditions is implemented, cost associated with inefficient recruitments (either as a reaction against the threat of intervention, or to comply with a representation quota) –, they may be optimally combined depending on the planner's constraints.

Let us conclude by evoking some of the (many) areas that would benefit from future research and in which we believe our model could be useful.

Theoretical research: (a) Heterogeneous time horizons. Heterogeneous time horizons may stem from different preferences, different positions within the organization's hierarchy, or different outside options (due e.g., to heterogeneous sectorial or geographical mobility). Members' heterogeneous time horizons affect their willingness to invest for the future. As we showed (see Proposition 2.iv), discrimination against the minority is an investment benefitting patient majority members. Would "older" members (i.e. with a shorter time horizon) be more meritocratic than "younger" members, and thus should they be allocated more power over recruitments? Relatedly, in a hierarchy with control located at the senior level, would the glass ceiling be shattered by a shorter time horizon for seniors (e.g., stemming from an internal rule mandating that seniors only serve as such for a short duration)?

(b) Integrity of quality assessment exercises. One of our insights on the policy side is that quality assessment exercises promote meritocracy 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 only because they suffer from some innate trait that is unrelated to quality (gender, ethnicity…), but also because they are perceived as lower-quality agents by the majority group. Mandating diversity in the assessment panels may then be less effective than when differentiating (in/out-group) traits are perceived to be horizontal. Capturing this may require a diversity of perhaps-motivated beliefs as to what constitutes high-quality work, and would shed light on how science progresses.

(c) 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.

(d) Dismissals. We have assumed that the organization's members have tenure. While this is a reasonable approximation in a number of coopting organizations (academia, cooperatives, trade or monetary unions, academies, some civil service and/or judiciary positions in some countries[47]), one could study the opposite polar case in which members have short-term appointments that need to be renewed over time. The preferences of talented and untalented members of a given horizontal group are then not perfectly aligned. The latter's position is more at risk in case of a change of majority, as the new majority can always replace them and bring in its own untalented members (recall that talent is the scarce factor). Hence, talented and untalented majority members may have diverging opinions on meritocracy, with the tal-

[47]In some countries (e.g., in France and in Spain), recruitment into some civil-service corps – which offer permanent positions – is made via committees composed of incumbent members of the corps. See for instance Bagues and Esteve-Volart (2010) on recruitments into the main corps of the Spanish judiciary.

ented opting for meritocracy while untalented ones strongly prefer entrenchment. One could even envision cases in which the untalented majority members would vote for one of their clones over a talented majority candidate, giving rise to a new, vertical form of entrenchment. We conjecture that meritocracy might be best promoted by awarding tenure and thereby securing the positions of the less talented in case of a change in majority.

Empirical investigations. 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 (e.g., political polarization) and decreased in others (as when social norms penalize a lack of diversity). Depending on factors such as initial conditions, the nature of internal interactions, the size of the organization[48] or the competitiveness of the market for talent, this evolution should impact dependent variables such as the quality of recruitments and the heterogeneity within and across organizations. Does patience (e.g., longer-term perspectives for members within the organization) foster entrenchment as the model predicts? For example, the model's predictions on the role of patience may be particularly relevant when applied to local communities. People with low prospects of ever leaving a region or a neighborhood (the "somewheres", to borrow from Goodhart 2017), should be expected to be more inclined to entrench themselves, i.e. be opposed to a large immigration that would make them become a minority, while by contrast, highly mobile individuals (the "anywheres") should be more tolerant/less sensitive. In addition, the model's predictions on the impact of policy interventions could be tested: How do organizations react in practice, and do policies backfire as predicted? The model's results regarding the collegial bias against meritocracy and its drivers could also be tested. For instance, do uncertain voting participation and imperfect group allegiance trigger super-entrenchment?[49] Does homogamic evaluation capability threaten meritocracy and harm welfare?[50] We leave these empirical questions as well as the theoretical ones to future investigation.

References

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

[48] Our model indeed predicts that at least for δ_0 small, small organizations are more likely to be in the meritocratic equilibrium, yet it also predicts that conditional on entrenchment, larger organizations make on average more meritocratic recruitments.

[49] In particular, other things being equal, do lower levels of trust within groups lead to higher levels of entrenchment as the model suggests? Conversely, do more tightly-knit groups foster meritocracy?

[50] As the polarization of a society increases, reciprocal knowledge across groups may be expected to recede, hence generating the asymmetric information structure of Section 4.2.1. As a consequence, polarization in a society may jeopardize meritocracy both by raising homophily benefits, and by lowering information on out-group individuals.

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". *Quarterly Journal of Economics*, 115(4):1167–1199.

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". *Review of Economic Studies*, 77(4):1301–1328.

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.

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". *Review of Economic Studies*, 86(1):191–239.

Board, S., Meyer-ter-Vehn, M., and Sadzik, T. (2020). "Recruiting Talent". R&R *Review of Economic Studies.*

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.

Casella, A., Palfrey, T., and Riezman, R. (2008). "Minorities and Storable Votes". *Quarterly Journal of Political Science*, 3(2):165–200.

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

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

Fershtman, D. and Pavan, A. (2021). ""Soft" Affirmative Action and Minority Recruitment". *American Economic Review: Insights*, 3(1):1–18.

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.

Goodhart, D. (2017). *The Road to Somewhere: The Populist Revolt and the Future of Politics*. C. Hurst & Co.

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

Jackson, M. O. and Sonnenschein, H. F. (2007). "Overcoming Incentive Constraints by Linking Decisions". *Econometrica*, 75(1):241–257.

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

Lipnowski, E. and Ramos, J. (2020). "Repeated Delegation". *Journal of Economic Theory*, 188(C).

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

Moisson, P.-H. and Tirole, J. (2024). "Managing Diversity to Attract Talent". mimeo.

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

Moreira, D. and Pérez, S. (2022). "Civil Service Exams and Organizational Performance: Evidence from the Pendleton Act". Conditionally accepted at *American Economic Journal: Applied Economics*.

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)". *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.

Online Appendix

A Proof of Lemma 1

A.1 Value functions for majority and minority members with the canonical strategies

Let us consider the canonical strategies:

(i) Members of the majority (all) vote for the majority candidate if the latter is at least as talented as the minority candidate.

(ii) When the minority candidate is more talented, members of a type-X majority (all) vote for the majority candidate with probabilities $\{\sigma_X(M)\}_{M\in\{k,\ldots,N-1\}}$ with $\sigma_X(k)\in\{0,1\}$ and $\sigma_X(M)=0$ if $M>k$.

Value function for a majority member. Let $V^-_{i,X}$ denote the expected value function conditional on the minority candidate being more talented, and $V^+_{i,X}$ denote the expected value function conditional on the complementary event. The value function for a majority member writes for any $k\le M\le N-1$,[51],

$$V_{M,X}=xV^-_{M,X}+(1-x)V^+_{M,X} \tag{3}$$

$$\text{where}\quad\begin{cases} V^-_{M,X}=\sigma_X(M)\Big[b_X+\delta\Big(\dfrac{M}{N-1}V_{M,X}+(1-\dfrac{M}{N-1})V_{M+1,X}\Big)\Big] \\ \qquad\qquad +(1-\sigma_X(M))\Big[s+\delta\Big(\dfrac{M-1}{N-1}V_{M-1,X}+(1-\dfrac{M-1}{N-1})V_{M,X}\Big)\Big] \\ V^+_{M,X}=b_X+\dfrac{\overline{x}}{1-x}s+\delta\Big(\dfrac{M}{N-1}V_{M,X}+\Big(1-\dfrac{M}{N-1}\Big)V_{M+1,X}\Big)\end{cases}$$

With probability x, the (type-X) majority faces a trade-off between selecting a talented minority member (yielding payoff s) and picking the less talented majority candidate (yielding payoff b_X). With probability $1-x$, the majority candidate is at least as talented as the minority one, and the majority candidate brings average payoff $b_X+\overline{x}s/(1-x)$, where $\overline{x}/(1-x)$ is the conditional probability of that candidate's being talented. Recruiting a majority candidate when the majority has size M in period t yields an end-of-period majority size of $M+1$. From the perspective of a majority member, 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 embedded

[51] Equation (3) applies even when $M=N-1$ as the majority size $M+1$ becomes irrelevant (its probability being nil).

in the discount factor $\delta \equiv \delta_0(1-1/N)$), 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 $M+1$. The value function for a (type-X) minority member writes for any $k \leq M \leq N-2$:

$$V_{N-M-1,X} = xV^-_{N-M-1,X} + (1-x)V^+_{N-M-1,X} \tag{4}$$

$$\text{where } \begin{cases} V^-_{N-M-1,X} = \sigma_Y(M)\delta\Big(\dfrac{M+1}{N-1}V_{N-M-1,X} + (1-\dfrac{M+1}{N-1})V_{N-M-2,X}\Big) \\ \qquad\qquad +(1-\sigma_Y(M))\Big[s + b_X + \delta\Big(\dfrac{M}{N-1}V_{N-M,X} + (1-\dfrac{M}{N-1})V_{N-M-1,X}\Big)\Big] \\ V^+_{N-M-1,X} = \dfrac{\overline{x}}{1-x}s + \delta\Big(\dfrac{M+1}{N-1}V_{N-M-1,X} + \Big(1-\dfrac{M+1}{N-1}\Big)V_{N-M-2,X}\Big) \end{cases}$$

A.2 Continuation values with the canonical strategies

Let us begin with a useful result, which we will use repeatedly throughout our analysis. We index the canonical strategies by $r \in \{m, e\}$, where m stands for the canonical meritocratic strategy and e for the canonical entrenchment one. To alleviate the notation, we omit the subscript $X \in \{A, B\}$ as we restrict our attention to a single group.

Lemma A.1. ***(Majority continuation values with the canonical strategies)*** *Fix $V_{k-1} \in \mathbb{R}$ (continuation value upon losing control) and consider the value function $(V^r_M(V_{k-1}))_{M\geq k}$ associated with the canonical strategy $r \in \{m, e\}$ given V_{k-1}. Then,*

(i) For $r = e$, the continuation value $V^e_M(V_{k-1})$ increases with $M \geq k$ and has decreasing differences (i.e., $u^e_M(V_{k-1}) \equiv V^e_{M+1}(V_{k-1}) - V^e_M(V_{k-1})$ decreases with $M \geq k$), strictly so if $x > 0$.

(ii) For $r = m$, if $V^m_k(V_{k-1}) \geq V_{k-1}$, the continuation value $V^m_M(V_{k-1})$ increases with $M \geq k$ and has decreasing differences (i.e., $u^m_M(V_{k-1}) \equiv V^m_{M+1}(V_{k-1}) - V^m_M(V_{k-1})$ decreases with $M \geq k$), strictly so if $V^m_k(V_{k-1}) > V_{k-1}$.

Proof. Let $r \in \{m, e\}$. By definition of the canonical strategies, for any $M \in \{k+1, ..., N-1\}$,

$$V_M^r(V_{k-1}) = (\overline{x} + x)s + (1-x)b + (1-x)\delta\Big[\frac{M}{N-1}V_M^r(V_{k-1}) + \Big(1 - \frac{M}{N-1}\Big)V_{M+1}^r(V_{k-1})\Big] + x\delta\Big[\frac{M-1}{N-1}V_{M-1}^r(V_{k-1}) + \Big(1 - \frac{M-1}{N-1}\Big)V_M^r(V_{k-1})\Big].$$

For $M = k$, the same recursive equation holds for the canonical meritocratic strategy ($r = m$), while for the canonical entrenchment strategy ($r = e$),

$$V_k^e(V_{k-1}) = \overline{x}s + b + \delta\Big[\frac{k}{N-1}V_k^e(V_{k-1}) + \Big(1 - \frac{k}{N-1}\Big)V_{k+1}^e(V_{k-1})\Big].$$

Consequently, for any $M \in \{k+1, ..., N-1\}$, letting $u_M^r(V_{k-1}) \equiv V_{M+1}^r(V_{k-1}) - V_M^r(V_{k-1})$,

$$\Big[1 - \delta x\Big(1 - \frac{M}{N-1}\Big) - \delta(1-x)\frac{M}{N-1}\Big]u_M^r(V_{k-1}) = \delta x\frac{M-1}{N-1}u_{M-1}^r(V_{k-1}) + \delta(1-x)\Big(1 - \frac{M+1}{N-1}\Big)u_{M+1(V_{k-1})}^r. \tag{5}$$

Moreover, for $M = k$, the canonical meritocratic strategy (still) yields (5), i.e.

$$\Big[1 - \delta x\Big(1 - \frac{k}{N-1}\Big) - \delta(1-x)\frac{k}{N-1}\Big]u_k^m(V_{k-1}) = \delta x\frac{k-1}{N-1}u_{k-1}^m(V_{k-1}) + \delta(1-x)\Big(1 - \frac{k+1}{N-1}\Big)u_{k+1}^m(V_{k-1}).$$

whereas the canonical entrenchment strategy yields

$$u_k^e(V_{k-1}) = x(s-b) + \delta(1-x)\Big[\Big(1 - \frac{k+1}{N-1}\Big)u_{k+1}^e(V_{k-1}) + \frac{k}{N-1}u_k^e(V_{k-1})\Big]. \tag{6}$$

We show the result by contradiction and by induction. Suppose $u_{N-2}^r(V_{k-1}) \leq 0$. Then, Equation (5) for $M = N-2$ implies

$$\Big[1 - \delta\frac{x}{N-1} - \delta(1-x)\frac{N-2}{N-1}\Big]u_{N-2}^r(V_{k-1}) = \delta x\frac{N-3}{N-1}u_{N-3}^r(V_{k-1})$$

Therefore, $u_{N-3}^r(V_{k-1}) \leq 0$ and $u_{N-3}^r(V_{k-1}) \leq u_{N-2}^{(}V_{k-1})r$. We then proceed by induction to show that for any $M \in \{k+1, ..., N-2\}$, $u_{M-1}^r(V_{k-1}) \leq u_M^r(V_{k-1}) \leq 0$. Suppose the result holds for all indices in $\{M+1, ..., N-2\}$. Then, (5) implies

$$\Big[1 - \delta x\Big(1 - \frac{M}{N-1}\Big) - \delta(1-x)\frac{M}{N-1}\Big]u_M^r(V_{k-1}) \geq \delta x\frac{M-1}{N-1}u_{M-1}^r(V_{k-1}) + \delta(1-x)\Big(1 - \frac{M+1}{N-1}\Big)u_M^r(V_{k-1})$$

i.e.,

$$\left[1-\delta x\left(1-\frac{M}{N-1}\right)-\delta(1-x)\frac{N-2}{N-1}\right]u^r_M(V_{k-1})\geq\delta x\frac{M-1}{N-1}u^r_{M-1}(V_{k-1})$$

Consequently, $u^r_{M-1}(V_{k-1})\leq u^r_M(V_{k-1})\leq 0$. The result follows by induction. In particular, one has $u^r_k(V_{k-1})\leq u^r_{k+1}(V_{k-1})\leq 0$.

However, consider the canonical entrenchment strategy and suppose $x>0$ (the case $x=0$ is analogous). (6) then implies that

$$0\geq\left[1-\delta(1-x)\frac{N-2}{N-1}\right]u^e_k(V_{k-1})\geq x(s-b)>0,$$

which is a contradiction. Similarly, consider the canonical meritocratic strategy and suppose that $V^m_k(V_{k-1})>V_{k-1}$ (the weak inequality case is analogous). Using (5) in $M=k$ allows to extend the induction argument to show that $u^m_{N-2}\leq 0$ implies $u^m_{k-1}(V_{k-1})\leq 0$, i.e. $V^m_k(V_{k-1})\leq V_{k-1}$, which is a contradiction.

Therefore, for any $r\in\{m,e\}$, $u^r_{N-2}(V_{k-1})<0$. Using (5), one then has by induction that

$$u^r_k(V_{k-1})>...>u^r_{N-2}(V_{k-1})>0.$$

as was to be shown. □

A.3 Proof of Lemma 1

Let v (resp. w) denote the incremental value brought to a member of the majority by the minority (resp. majority) candidate. So $v\in\{0,s\}$, $w\in\{b,b+s\}$, and $v>w$ if and only if $(v,w)=(s,b)$ (otherwise $v<w$). Throughout the Online Appendix, we refer to the incremental value brought by current-period hires as a "flow payoff" (slighltу abusing vocabulary as this incremental value captures the discounted sum of present and future quality and homophily benefits, if any).

Let $\mathcal{C}\equiv[0,((\overline{x}+x)s+(1-x)b)/(1-\delta)]$. All vectors of value functions $(V_k,..,V_{N-1})$ necessarily belong to $\mathcal{C}^k$ as for any $s\geq b$, $\mathbb{E}_{v,w}[\max(v,w)]=(\overline{x}+x)s+(1-x)b$. By construction, given any $V_{k-1}\in\mathcal{C}$, the majority faces an optimal control problem, and there exists a unique sequence of majority value functions $(V_k(V_{k-1}),...,V_{N-1}(V_{k-1}))$ solving the Bellman equations:

$$\forall i\geq k,\qquad V_i=\mathbb{E}_{v,w}\Bigg[\max\Bigg\{v+\delta\Big(\frac{i-1}{N-1}V_{i-1}+\Big(1-\frac{i-1}{N-1}\Big)V_i\Big),\\ w+\delta\Big(\frac{i}{N-1}V_i+\Big(1-\frac{i}{N-1}\Big)V_{i+1}\Big)\Bigg\}\Bigg]$$

Hence, rewriting (1), the majority's choice at size M between two candidates with profiles (v, w) is determined by the following comparison:

$$v - w + \delta\Big[\frac{M-1}{N-1}(V_M - V_{M-1}) + \Big(1 - \frac{M}{N-1}\Big)(V_{M+1} - V_M)\Big] \lesseqgtr 0. \qquad (7)$$

Given any $V_{k-1} \in \mathcal{C}$, the majority can always guarantee a sequence of value functions such that $V_M > V_{k-1}$ for any $M \geq k$, for instance by following the canonical meritocratic strategy (making meritocratic recruitments at all majority sizes) as such a strategy yields a flow payoff equal to $\mathbb{E}_{v,w}[\max(v, w)] = (\overline{x} + x)s + (1 - x)b \geq (1 - \delta)V_{k-1}$ at all majority sizes.

Hence in particular, the solution to the Bellman equations given V_{k-1} satisfies $V_{k+1}(V_{k-1}) > V_{k-1}$, and thus for $M = k$, (7) writes as

$$v - w - \delta\frac{k-1}{N-1}(V_{k+1}(V_{k-1}) - V_{k-1}) \leq v - w.$$

Hence, it is never optimal for a majority with size k to recruit the minority candidate whenever $v < w$ (i.e. whenever the majority candidate is at least as talented as the minority one).

Fix $V_{k-1} \in \mathcal{C}$. Let us show that for any $V_{k-1} \in \mathcal{C}$, the majority's best response among pure Markov Perfect strategy is either canonical meritocracy or canonical entrenchment.

Consider the sequence of value functions $(V^{\mathrm{e}}_M(V_{k-1}))_{M\geq k}$ generated by the canonical entrenchment strategy given V_{k-1}: the sequence $(V^{\mathrm{e}}_M(V_{k-1}))_{M\geq k}$ is defined recursively by (5)-(6), i.e. satisfies

$$V^e_k(V_{k-1}) = \mathbb{E}[w] + \delta\Big[\frac{k}{N-1}V^e_k(V_{k-1}) + \frac{k-1}{N-1}V^e_{k+1}(V_{k-1})\Big]$$

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

$$\begin{aligned} V^{\mathrm{e}}_M(V_{k-1}) = \mathbb{E}_{v,w}[\max(v, w)] &+ \delta x\Big[\frac{M-1}{N-1}V^{\mathrm{e}}_{M-1}(V_{k-1}) + \frac{N-M}{N-1}V^{\mathrm{e}}_M(V_{k-1})\Big] \\ &+ \delta(1 - x)\Big[\frac{M}{N-1}V^{\mathrm{e}}_M(V_{k-1}) + \frac{N-M-1}{N-1}V^{\mathrm{e}}_{M+1}(V_{k-1})\Big] \end{aligned}$$

Let us distinguish three cases, depending on whether $s - b - \delta\frac{k-1}{N-1}(V^e_{k+1}(V_{k-1}) - V_{k-1})$ is strictly negative, strictly positive, or nil.

Case 1. Suppose the following inequality holds:

$$s - b - \delta\frac{k-1}{N-1}(V^e_{k+1}(V_{k-1}) - V_{k-1}) < 0. \qquad (8)$$

Let us show that the sequence of majority value functions $(V^{\mathrm{e}}_M(V_{k-1}))_{M\geq k}$ solves the Bellman equations given V_{k-1}. By Lemma A.1, $u^{\mathrm{e}}_M \equiv V^{\mathrm{e}}_{M+1} - V^{\mathrm{e}}_M > 0$ and $u^{\mathrm{e}}_{M+1} \leq u^{\mathrm{e}}_M$ for all $M \geq k$. Hence, (7) implies that given the continuation values induced by the canonical entrenchment strategy, it is strictly optimal for the majority at any majority size $M \geq k$ to recruit its in-group candidate whenever he is at least as talented as the minority one (as then $v > w$).

Moreover, (8) and (7) imply that given the continuation values induced by the canonical entrenchment strategy, it is optimal for the majority at size $M = k$ to recruit the majority candidate even when he is less talented than the minority candidate. In addition, (6) together with the inequality $u^{\mathrm{e}}_{k+1} \leq u^{\mathrm{e}}_k$ imply that

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

and therefore, using again (6),

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

where the second inequality follows from $\delta < (N-1)/N$. Hence, by monotonicity of the sequence $(u^{\mathrm{e}}_M)_{M \geq k}$, for any majority size $M \geq k+1$, (7) implies that, given the continuation values induced by the canonical entrenchment strategy, it is strictly optimal for the majority to recruit the minority candidate whenever she is more talented than the majority candidate.

Therefore, the sequence of majority value functions $(V^{\mathrm{e}}_M(V_{k-1}))_{M \geq k}$ solves the Bellman equations given V_{k-1}, and as the latter have a unique solution, $V_M(V_{k-1}) = V^{\mathrm{e}}_M(V_{k-1})$ for any $M \geq k$. Identifying the strategies from the value functions (using (7)), if (8) holds, the majority's best response to V_{k-1} among pure Markov Perfect strategies is thus the canonical entrenchment strategy.

Case 2. Suppose the following inequality holds:

$$s - b - \delta\frac{k-1}{N-1}\left(V^{e}_{k+1}(V_{k-1}) - V_{k-1}\right) > 0, \tag{9}$$

To alleviate the notation, let

$$\Delta \equiv s - b - \delta\frac{k-1}{N-1}\left(V^{e}_{k+1}(V_{k-1}) - V_{k-1}\right).$$

Consider the sequence of value functions $(V^{\mathrm{m}}_M(V_{k-1}))_{M \geq k}$ generated by the canonical meritocratic strategy given V_{k-1}: the sequence $(V^{\mathrm{m}}_M(V_{k-1}))_{M \geq k}$ is defined recursively by (5) for any

$M \geq k$, i.e. satisfies for all $M \geq k$

$$V_M^{\mathrm{m}}(V_{k-1}) = \mathbb{E}_{v,w}[\max(v,w)] + \delta x\Big[\frac{M-1}{N-1}V_{M-1}^{\mathrm{m}}(V_{k-1}) + \frac{N-M}{N-1}V_M^{\mathrm{m}}(V_{k-1})\Big] + \delta(1-x)\Big[\frac{M}{N-1}V_M^{\mathrm{m}}(V_{k-1}) + \frac{N-M-1}{N-1}V_{M+1}^{\mathrm{m}}(V_{k-1})\Big]$$

with $V_{k-1}^{\mathrm{m}}(V_{k-1}) \equiv V_{k-1}$.

Then, using the recursive expressions of the continuation values induced by the canonical meritocratic and canonical entrenchment strategies, for any $M \geq k+1$,

$$\begin{aligned} V_M^{\mathrm{m}}(V_{k-1}) - V_M^{\mathrm{e}}(V_{k-1}) \\ = \delta x\Big[\frac{M-1}{N-1}(V_{M-1}^{\mathrm{m}}(V_{k-1}) - V_{M-1}^{\mathrm{e}}(V_{k-1})) + \frac{N-M}{N-1}(V_M^{\mathrm{m}}(V_{k-1}) - V_M^{\mathrm{e}}(V_{k-1}))\Big] \\ + \delta(1-x)\Big[\frac{M}{N-1}(V_M^{\mathrm{m}}(V_{k-1}) - V_M^{\mathrm{e}}(V_{k-1})) + \frac{N-M-1}{N-1}(V_{M+1}^{\mathrm{m}}(V_{k-1}) - V_{M+1}^{\mathrm{e}}(V_{k-1}))\Big] \end{aligned}$$

and for $M = k$,

$$\begin{aligned} V_k^{\mathrm{m}}(V_{k-1}) - V_k^{\mathrm{e}}(V_{k-1}) \\ = \Delta + \delta x\frac{k}{N-1}(V_k^{\mathrm{m}}(V_{k-1}) - V_k^{\mathrm{e}}(V_{k-1}))\Big] \\ + \delta(1-x)\Big[\frac{k}{N-1}(V_k^{\mathrm{m}}(V_{k-1}) - V_k^{\mathrm{e}}(V_{k-1})) + \frac{k-1}{N-1}(V_{k+1}^{\mathrm{m}}(V_{k-1}) - V_{k+1}^{\mathrm{e}}(V_{k-1}))\Big]. \end{aligned}$$

Using the recursive expressions of the $(V_M^m - V_M^e)_M$ yields that[52]

$$V_k^{\mathrm{m}}(V_{k-1}) - V_k^{\mathrm{e}}(V_{k-1}) > \ldots > V_{N-1}^{\mathrm{m}}(V_{k-1}) - V_{N-1}^{\mathrm{e}}(V_{k-1}) > 0.$$

As a consequence, by the recursive expression of $V_{k+1}^{\mathrm{m}}(V_{k-1}) - V_{k+1}^{\mathrm{e}}(V_{k-1})$,

$$\Big[1 - \delta(1-x) - \delta x\frac{k-1}{N-1}\Big](V_{k+1}^{\mathrm{m}}(V_{k-1}) - V_{k+1}^{\mathrm{e}}(V_{k-1})) < \delta x\frac{k}{N-1}(V_k^{\mathrm{m}}(V_{k-1}) - V_k^{\mathrm{e}}(V_{k-1})),$$

and thus by the recursive expression of $V_k^{\mathrm{m}}(V_{k-1}) - V_k^{\mathrm{e}}(V_{k-1})$,

$$\Big[1 - \delta(1-x) - \delta x\frac{k-1}{N-1}\Big](V_{k+1}^{\mathrm{m}}(V_{k-1}) - V_{k+1}^{\mathrm{e}}(V_{k-1})) < \frac{\delta x\frac{k}{N-1}}{1-\delta(1-x) - \delta x\frac{k}{N-1}}x\Delta.$$

Therefore,

$$\delta\frac{k-1}{N-1}(V_{k+1}^{\mathrm{m}}(V_{k-1}) - V_{k+1}^{\mathrm{e}}(V_{k-1})) < \Delta,$$

[52] One may for instance proceed as in the proof of Lemma A.1 and suppose by contradiction that $V_{N-1}^{\mathrm{m}}(V_{k-1}) - V_{N-1}^{\mathrm{e}}(V_{k-1}) \leq 0$.

and hence, by definition of Δ,

$$s - b - \delta \frac{k-1}{N-1} (V_{k+1}^m(V_{k-1}) - V_{k-1}) > 0. \tag{10}$$

Consequently, (10) and (7) imply that given the continuation values induced by the canonical meritocratic strategy, it is optimal for the majority at size $M = k$ to recruit the minority candidate whenever she is more talented than the majority candidate.

By construction, $V_{k-1} \in \mathcal{C}$, and thus $V_{k-1} \leq \mathbb{E}_{v,w}[\max(v,w)]/(1-\delta)$. Therefore, $V_k^m(V_{k-1}) \geq V_{k-1}$. Consequently, by Lemma A.1, $u_{k-1}^m \geq u_k^m \geq ... \geq u_{N-2}^m \geq 0$. Hence, (7) implies that given the continuation values induced by the canonical meritocratic strategy, it is indeed strictly optimal at any majority size $M \geq k$ for the majority to recruit the majority candidate whenever he is at least a talented as the minority candidate (as then $v > w$). Moreover, (10), the monotonicity of the sequence $(u_M^m)_{M \geq k-1}$, and (7) imply that given the continuation values induced by the canonical meritocratic strategy, it is strictly optimal at any majority size $M \geq k$ for the majority to recruit the minority candidate whenever she is more talented than the minority candidate.

Therefore, the sequence of value functions $(V_M^{\mathrm{m}}(V_{k-1}))_{M \geq k}$ generated by the canonical meritocratic strategy given V_{k-1} solves the Bellman equations, and as the latter have a unique solution, $V_i(V_{k-1}) = V_i^{\mathrm{m}}(V_{k-1})$ for any $i \geq k$. Identifying the strategies from the value functions (using (7)), if (9) holds, the majority's best response to V_{k-1} among pure Markov Perfect strategies is thus the canonical meritocratic strategy.

Case 3: Suppose that the following equality holds:

$$s - b - \delta \frac{k-1}{N-1} (V_{k+1}^e(V_{k-1}) - V_{k-1}) = 0,$$

i.e. the majority is indifferent between $\sigma(k) = 0$ and $\sigma(k) = 1$. Then, the above arguments imply that the sequences of value functions induced by the canonical entrenchment and canonical meritocratic strategies both solve the Bellman equations. Identifying the strategies from the value functions (using (7)), the majority has then two best responses (yielding the same continuation values): canonical meritocracy and canonical entrenchment.

B Proof of Lemma 2

The result for $N = 4$ derives from straightforward computations.[53] We assume in the following that $N \geq 6$.

Proof of (i). Consider first the canonical entrenchment strategies. For any $M \in \{k-1, ..., N-2\}$, let V_M^e denote the continuation value function with the canonical entrenchment strategy for both group, and let $u_M^e \equiv V_{M+1}^e - V_M^e$. As argued in the proof of Lemma A.1 (see Online Appendix A.2), the recursive expressions of the continuation value function yield for any $M \in \{k+1, ..., N-2\}$,

$$\left[1 - \delta x\left(1 - \frac{M}{N-1}\right) - \delta(1-x)\frac{M}{N-1}\right]u_M^e = \delta x \frac{M-1}{N-1}u_{M-1}^e + \delta(1-x)\left(1 - \frac{M+1}{N-1}\right)u_{M+1}^e, \tag{5}$$

and for $M = k$,

$$u_k^e = x(s-b) + \delta(1-x)\left[\left(1 - \frac{k+1}{N-1}\right)u_{k+1}^e + \frac{k}{N-1}u_k^e\right] \tag{6}$$

Therefore, the result follows straightforwardly from claim (i) in Lemma A.1.

Consider now the canonical meritocratic strategies. Let let V_M^e denote the continuation value function with the canonical entrenchment strategy for both group, and let $u_i^m \equiv V_{i+1}^m - V_i^m$ for any $i \in \{1, ..., N-2\}$. By construction, Equation (5) holds for any $M \in \{k, ..., N-2\}$. We use the same argument as in the proof of Lemma A.1 (by contradiction and by induction).

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

Consider now u_1^m. Writing the recursive expression of the value function in $M \in \{k+1, ..., N-1\}$ (thus writing V_{N-M-1}^m as a function of V_{N-M-2}^m, V_{N-M-1}^m and V_{N-M}^m), and then substracting the expression in $N-M-1$ from the expression in $N-M$ (and rearranging)

[53] Using (3) and (4), the canonical entrenched strategies yield

$$\left[1 - \frac{2\delta}{3}(1-x)\right](V_3^{\text{e}} - V_2^{\text{e}}) = x(s-b)$$

and thus $V_1^{\text{e}} = \overline{x}s/(1-\delta) < (b + \overline{x}s)/(1-\delta) < V_2^{\text{e}} < V_3^{\text{e}}$. Similarly for the meritocratic equilibrium:

$$\begin{cases} \left[1 - \frac{x\delta}{3} - \frac{2\delta}{3}(1-x)\right](V_3^{\text{m}} - V_2^{\text{m}}) = \frac{x\delta}{3}(V_2^{\text{m}} - V_1^{\text{m}}) \\ \left[1 - \delta(1-x)\right](V_2^{\text{m}} - V_1^{\text{m}}) = (1-2x)b + \delta\frac{(1-x)}{3}(V_3^{\text{m}} - V_2^{\text{m}}) \end{cases}$$

and thus $V_1^{\text{m}} < V_2^{\text{m}} < V_3^{\text{m}}$, and $V_2^{\text{m}} - V_1^{\text{m}} > V_3^{\text{m}} - V_1^{\text{m}}$.

yields for any $M \in \{k+2, ..., N-2\}$:

$$\left[1 - \delta(1-x)\frac{M}{N-1} - \delta x\left(1 - \frac{M}{N-1}\right)\right]u^m_{N-M-1}$$
$$= \delta(1-x)\left(1 - \frac{M+1}{N-1}\right)u^m_{N-M-2} + \delta x\frac{M-1}{N-1}u^m_{N-M} \quad (11)$$

and in particular,

$$\left[1 - \delta\frac{x}{N-1} - \delta(1-x)\frac{N-2}{N-1}\right]u^m_1 = \delta x\frac{N-3}{N-1}u^m_2$$

By the usual induction argument using (11), $u^m_1 > 0$ implies $0 < u^m_1 < u^m_2 < ... < u^m_{k-2} < u^m_{k-1}$, which contradicts $u^m_{k-1} \leq 0$. Hence $u^m_1 \leq 0$ and the same induction argument now implies $0 \geq u^m_1 \geq u^m_2 \geq ... \geq u^m_{k-2} \geq u^m_{k-1}$.

However, substracting Equation (3) in k and Equation (4) in $k-1$ yields after rearranging:

$$\left[1 - \delta(1-x)\right]u^m_{k-1} = (1-2x)b + \delta(1-x)\left[\left(1 - \frac{k}{N-1}\right)u^m_k + \left(1 - \frac{k+1}{N-1}\right)u^m_{k-2}\right] \quad (12)$$

The contradiction then obtains by summing the above equation together with Equations (5) and (11) over all indices $i \in \{1, ..., N-2\}$ (and rearranging), which gives:

$$\left(1 - \delta\frac{x}{N-1} - \delta(1-x)\right)(u^m_1 + u^m_{N-2}) + (1-\delta)\sum_{i=2}^{N-3} u^m_i = (1-2x)b > 0$$

If $x < 1/2$, this contradicts the fact that $u^m_i \leq 0$ for all $i \in \{1, ..., N-2\}$. Therefore, $u^m_{N-2} > 0$. By induction, Equation (5) then implies that $0 < u^m_{N-2} < ... < u^m_{k-1}$.[54]

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

Claim *(iii)* again 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 $V^e_i - V^e_{N-1-i}$ for $i \in \{k, ..., N-2\}$ and using (3)-(4).[55]

Suppose by contradiction that $V^e_k - V^e_{k-1} \leq 0$. This implies that $V^e_{k+1} - V^e_{k-2} < V^e_k - V^e_{k-1} \leq$

[54] If $x = 1/2$, the same argument yields that $V^m_i = V^m_{i+1}$ for all i.

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

$$\left[1 - \delta(1-x)\frac{M}{N-1} - \delta x\left(1 - \frac{M-1}{N-1}\right)\right](V^e_M - V^e_{N-M-1}) - (1-2x)b + \frac{\delta}{N-1}\left[(1-x)u^e_{N-M-2} + xu^e_{N-M-1}\right]$$
$$= \delta(1-x)\left(1 - \frac{M}{N-1}\right)(V^e_{M+1} - V^e_{N-M-2}) + \delta x\frac{M-1}{N-1}(V^e_{M-1} - V^e_{N-M})$$

0, and thus by induction that $V^e_{N-1} - V^e_1 < V^e_{N-2} - V^e_1 < \ldots < V^e_k - V^e_{k-1} \leq 0$, which contradicts $V^e_{N-1} \geq V^e_{N-2}$ as shown above. (Another contradiction would be reached by summing as above the analogues of (5)-(11) and noting that the RHS is positive whenever $x \leq 1/2$.) Hence, $V^e_k - V^e_{k-1} > 0$. If $V^e_{k+1} - V^e_{k-2} \leq 0$, the same contradiction is reached again as then $V^e_{N-1} - V^e_1 < V^e_{N-2} - V^e_1 < \ldots < V^e_{k+1} - V^e_{k-2} \leq 0$ (Again, one could sum over $i \in \{k+1, ..., N-2\}$ the analogues of (5)-(11) and note that the RHS is positive whenever $x \leq 1/2$.) The result obtains by induction: for any $i \in \{k, ..., N-2\}$, $V^e_i - V^e_{N-1-i} > 0$.

C Proof of Proposition 2

Lemma 1 implies that in any pure-strategy Markov Perfect equilibrium (if any), each group plays a canonical strategy. Proposition 1 then implies that in the symmetric case ($b_A = b_B = b$), both groups must play the same (canonical) strategy, i.e. any pure-strategy Markov Perfect equilibrium (if any) is *canonical* and *symmetric*. This establishes claim (i).

Let us now study the existence regions of the canonical equilibria.

C.1 A necessary and sufficient condition for existence

Lemma C.1. *There exists no profitable one-shot deviation from a canonical strategy at any majority size and for any realization of the candidates' vertical types if and only if there exists no profitable deviation when $M = k$ and the minority candidate is strictly more talented.*

Proof. We know by Lemma 1 that a group's best response when it has the majority (group size $i \geq k$) is a canonical strategy – either canonical meritocracy or canonical entrenchment. The two canonical strategies coincide at all majority sizes and all profiles of current-period candidates, except at majority size $M = k$ when the minority candidate is strictly more talented than the majority candidate. The result follows. □

C.2 Existence regions

Let us introduce the notation for transition probabilities for group sizes *from the perspective of (in- or out-) group members*: for any horizontal group within the organization, we refer to the transition probability from group sizes i to j *from an (in- or out-) group member's perspective* as the probability that the group's size goes from i to j *conditional on the given*

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

$$\left[1 - \delta\frac{k}{N-1}\right](V^e_k - V^e_{k-1}) = b - \frac{\delta}{N-1}u^e_{k-2} + \delta\left(1 - \frac{k}{N-1}\right)(V^e_{k+1} - V^e_{k-2}),$$

$$\left[1 - \delta(1-x)\frac{N-2}{N-1} - \delta x\frac{2}{N-1}\right](V^e_{N-2} - V^e_1) = (1-2x)b - \frac{\delta x}{N-1}u_1 + \delta\frac{(1-x)}{N-1}(V^e_{N-1} - V^e_1) + \delta x\frac{N-3}{N-1}(V^e_{N-3} - V^e_2)$$

The result follows, as we know from above that in the entrenchment equilibrium, $u^e_i \leq 0$ for any $i \leq k-2$.

member being still a member of the organization then.

Namely, for regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$, let $p^{\mathrm{r}}_{i,j}$ be the one-period transition probability from an in-group member's perspective, i.e. the probability that a group size moves from $i \geq 1$ to $j \geq 1$ from one period to another conditional on the given group member still being in the organization in the following period (which has probability $(N-1)/N$). As an illustration, for any $M > k$ and in the entrenchment equilibrium ($\mathrm{r} = \mathrm{e}$), $p^{\mathrm{r}}_{i,j}$ is the probability from a majority member's perspective that the majority size moves from $i \geq k$ to $j \geq k$ from one period to another conditional on the majority member still being in the organization in the following period. Consequently,

$$\begin{cases} p^{\mathrm{e}}_{M,M+1} = (1-x)\Big(1-\dfrac{M+1}{N}\Big)\dfrac{N}{N-1} = (1-x)\Big(1-\dfrac{M}{N-1}\Big) \\ p^{\mathrm{e}}_{M,M} = \Big[(1-x)\dfrac{M}{N} + x\Big(1-\dfrac{M}{N}\Big)\Big]\dfrac{N}{N-1} = (1-x)\dfrac{M}{N-1} + x\Big(1-\dfrac{M-1}{N-1}\Big) \\ p^{\mathrm{e}}_{M,M-1} = x\dfrac{M-1}{N}\dfrac{N}{N-1} = x\dfrac{M-1}{N-1} \\ p^{\mathrm{e}}_{M,j} = 0 \qquad \text{if } |M-j| > 1. \end{cases} \tag{13}$$

and

$$\begin{cases} p^{\mathrm{e}}_{k,k+1} = \Big(1-\dfrac{k+1}{N}\Big)\dfrac{N}{N-1} = 1-\dfrac{k}{N-1} \\ p^{\mathrm{e}}_{k,k} = \dfrac{k}{N}\dfrac{N}{N-1} = \dfrac{k}{N-1} \\ p^{\mathrm{e}}_{k,k-1} = 0 \end{cases} \tag{14}$$

For any $i, j \in \{1, ..., N-1\}$ and $t \geq 0$, let $\pi^{\mathrm{r}}_{i,j}(t)$ be the t-period transition probability from i to j in regime r from an in-group member's perspective, i.e. the probability that starting from i, the group's size is equal to j after t periods conditional on the group member still being in the organization. Hence, for any $i, j \in \{1, ..., N-1\}$ and $t \geq 1$,

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

and $\pi^{\mathrm{r}}_{i,j}(1) = p^{\mathrm{r}}_{i,j}$.

Similarly, let $\hat{p}^{\mathrm{r}}_{i,j}$ be the transition probability from an *out-group* member's perspective, i.e. the probability that a group's size moves from $i \geq k$ to j from one period to another conditional on the other group member still being in the organization in the following period (which has probability $(N-1)/N$). As an illustration, for any $M > k$ and in the entrenched equilibrium, $\hat{p}^{\mathrm{e}}_{i,j}$ is the transition probability from a minority member's perspective, and thus

$$\begin{cases} \hat{p}^{\mathrm{e}}_{M,M+1} = (1-x)\Big(1-\dfrac{M+2}{N}\Big)\dfrac{N}{N-1} = (1-x)\Big(1-\dfrac{M+1}{N-1}\Big) \\ \hat{p}^{\mathrm{e}}_{M,M} = \Big[(1-x)\dfrac{M+1}{N} + x\Big(1-\dfrac{M+1}{N}\Big)\Big]\dfrac{N}{N-1} = (1-x)\dfrac{M+1}{N-1} + x\Big(1-\dfrac{M}{N-1}\Big) \\ \hat{p}^{\mathrm{e}}_{M,M-1} = x\dfrac{M}{N}\dfrac{N}{N-1} = x\dfrac{M}{N-1} \\ p^{\mathrm{e}}_{M,j} = 0 \qquad \text{if } |M-j|>1. \end{cases} \tag{15}$$

and

$$\begin{cases} \hat{p}^{\mathrm{e}}_{k,k+1} = \Big(1-\dfrac{k+2}{N}\Big)\dfrac{N}{N-1} = 1-\dfrac{k+1}{N-1} \\ \hat{p}^{\mathrm{e}}_{k,k} = \dfrac{k+1}{N}\dfrac{N}{N-1} = \dfrac{k+1}{N-1} \\ \hat{p}^{\mathrm{e}}_{k,k-1} = 0 \end{cases} \tag{16}$$

For any $i,j \in \{1,...,N-1\}$, and $t \geq 0$, let $\hat{\pi}^{\mathrm{r}}_{i,j}(t)$ be the t-period transition probability from i to j in regime r from an out-group member's perspective, i.e. the probability that starting from i, the group's size is equal to j after t periods conditional on the out-group member still being in the organization. Hence, for any $i,j \in \{1,...,N-1\}$ and $t \geq 1$,

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

and $\pi^{\mathrm{r}}_{i,j}(1) = p^{\mathrm{r}}_{i,j}$.

For the meritocratic equilibrium, transition probabilities are given by (13) for in-group members, and by (15) for out-group members at all group sizes ($i \in \{1,...,N-1\}$).

Note that because probabilities sum to 1,

$$\begin{cases} \Big(\displaystyle\sum_{i=k+1}^{N-1}\hat{\pi}^{\mathrm{e}}_{k,i}(t)\Big) - \Big(\displaystyle\sum_{i=k+1}^{N-1}\pi^{\mathrm{e}}_{k+1,i}(t)\Big) = -\Big(\hat{\pi}^{\mathrm{e}}_{k,k}(t) - \pi^{\mathrm{e}}_{k+1,k}(t)\Big) \\ \Big(\displaystyle\sum_{i=k}^{N-1}\pi^{\mathrm{m}}_{k+1,i}(t)\Big) - \Big(\displaystyle\sum_{i=k}^{N-1}\pi^{\mathrm{m}}_{k-1,i}(t)\Big) = -\Big[\Big(\displaystyle\sum_{i=1}^{k-1}\pi^{\mathrm{m}}_{k+1,i}(t)\Big) - \Big(\displaystyle\sum_{i=1}^{k-1}\pi^{\mathrm{m}}_{k-1,i}(t)\Big)\Big] \end{cases} \tag{17}$$

C.2.1 Proof of claims (ii) and (iii)

We now turn to the statement of the existence result. Building on Lemma C.1, let us examine the case in which the majority is tight ($M = k$) and the minority candidate is more talented.

Necessary and sufficient condition for existence of the meritocratic equilibrium. Leaving control considerations aside, 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 occurs 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 thus linked to homophily benefits when the two candidates equally talented (which has probability $1-2x$), as control allows one to select the in-group candidate. So, 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 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 C.1 implies that this condition is in fact also sufficient: as intuitive, deviations from meritocracy are less appealing further away from a tight majority size, i.e. from immediate control considerations.

Necessary and sufficient condition for existence of the entrenched equilibrium. Choosing the less talented majority candidate yields a direct payoff loss $s-b$. If the majority has size k, then with probability $(k-1)/(N-1)$, the surrendering of control translates into a permanent loss of homophily benefits whenever the two candidates are equally talented, which has probability $1-2x$. This cost is equal to

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

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 an 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

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

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

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

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

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

Let Inequality (18) be the inequality:

$$1+\delta\frac{k-1}{N-1}x\sum_{t=0}^{+\infty}\delta^t\Big(\pi^{\mathrm{e}}_{k+1,k}(t)-\hat{\pi}^{\mathrm{e}}_{k,k}(t)\Big)>0. \tag{18}$$

Define ρ^{e} as

$$\rho^{\mathrm{e}}\equiv\begin{cases}\dfrac{1+\dfrac{k-1}{N-1}\dfrac{\delta}{1-\delta}(1-2x)+\delta\dfrac{k-1}{N-1}x\sum_{t=0}^{+\infty}\delta^t\Big(\pi^{\mathrm{e}}_{k+1,k}(t)+\hat{\pi}^{\mathrm{e}}_{k,k}(t)\Big)}{1+\delta\dfrac{k-1}{N-1}x\sum_{t=0}^{+\infty}\delta^t\Big(\pi^{\mathrm{e}}_{k+1,k}(t)-\hat{\pi}^{\mathrm{e}}_{k,k}(t)\Big)} & \text{if (18) holds,}\\[2ex]+\infty & \text{otherwise.}\end{cases}$$

Hence, the entrenched equilibrium exists only if $s/b \leq \rho^{\mathrm{e}}$. As the series term in (18) is negative for all t (see Lemma C.2 below), there might exist an entrenched equilibrium for all values of s and b for δ sufficiently close to 1, and thus we set $\rho^{\text{e}} = +\infty$ if (18) fails. Nonetheless, we show that 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: $\rho^{\text{e}} < +\infty$ (see Section C.2.3 for the proof of this result).

Lemma C.1 yields that this necessary condition is also sufficient. Hence, the entrenched (resp. meritocratic) equilibrium exists if and only if $s/b \leq \rho^{\text{e}}$ (resp. $s/b \geq \rho^{\text{m}}$).

Lastly, we show that the cutoffs ρ^{e} and ρ^{m} satisfy the following inequalities:[56]

$$1 \leq 1 + \delta \frac{k-1}{N-1}(1-2x) \leq \rho^{\text{m}} \leq 1 + \frac{\delta}{1-\delta}\frac{k-1}{N-1}(1-2x) < \rho^{\text{e}} < +\infty \tag{19}$$

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.[57]

The bounds on ρ^{e} and ρ^{m} in Inequality (19) derive from the following lemma.

Lemma C.2. *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.[58]
(i) Define the process $M(t)$ (resp. $\hat{M}(t)$) as the probability distribution over majority sizes $\{k, ..., N-1\}$ from a majority (resp. minority) member's perspective. Hence, 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, $M(t) = (\pi^{\text{e}}_{k+1,k}(t), ..., \pi^{\text{e}}_{k+1,N-1}(t))$. 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, $\hat{M}(t) = (\hat{\pi}^{\text{e}}_{k,k}(t), ..., \hat{\pi}^{\text{e}}_{k,N-1}(t))$.

Let P (resp. $\hat{P}$) be the stochastic matrix associated with the process $M(t)$ (resp. $\hat{M}(t)$). As a consequence, for any $i, j \in \{1, ..., k\}$,

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

[56] The proof that $\rho^{\text{e}} < +\infty$ is delayed to Section C.2.3.

[57] 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 $(k-1)/(N-1)$).

[58] See, e.g., Kijima, M. (1997). "Monotone Markov Chains". In: *Markov Processes for Stochastic Modeling.* Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-3132-0_3.

We first note that for any $i > i'$ and any $j^* \in \{1, ..., k\}$, $\sum_{j \geq j^*} P_{ij} \geq \sum_{j \geq j^*} P_{i'j}$, i.e. $P_{i\cdot}$ *stochastically dominates* $P_{i'\cdot}$ whenever $i > i'$. Hence, P is *stochastically monotone*, and by the same argument, so is $\hat{P}$.

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 $\hat{M}(t)$, 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 $\hat{M}(t)$ (see for instance Theorem 3.31 in Kijima 1997).[59] 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 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 (19), 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[60] 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,

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

□

[59] A sketch of the proof is as follows. Proceed by induction on t. The result for $t = 0$ holds as the initial state $M(0) = (0, 1, 0, ...)$ stochastically dominates the initial state $\hat{M}(0) = (1, 0, , 0, ...)$. 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.

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

C.2.2 Proof of claim (iv)

The result derives from the explicit expressions of the existence thresholds together with Lemma C.2. Indeed, by Lemma C.2, for all $t \geq 0$,

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

Using term-by-term differentiation of the series yields the result: $\partial\rho^{\mathrm{m}}/\partial\delta_0 \geq 0, \partial\rho^{\mathrm{e}}/\partial\delta_0 \geq 0$ for all $\delta_0 \in [0,1)$. Moreover, using term-by-term differentiation of the series for ρ^{m} and explicit computations for ρ^{e} yields

$$\left.\frac{\partial\rho^{\mathrm{m}}}{\partial\delta_0}\right|_{\delta_0=0} = \frac{k-1}{N}(1-2x) \qquad \text{and} \qquad \left.\frac{\partial\rho^{\mathrm{e}}}{\partial\delta_0}\right|_{\delta_0=0} = \frac{k-1}{N}$$

Lastly, the explicit expressions of the existence thresholds yield that for δ_0 close to 0, ρ^{m} and ρ^{e} increase with the size of the organization $N = 2k$.

C.2.3 Entrenchment exists only on a finite interval ($\rho^{\mathrm{e}} < \infty$)

We show in this section that $\rho^{\mathrm{e}} < \infty$.[61] The result immediately follows from the explicit expression of ρ^{e} for $k = 2$. Hence, let $k \geq 3$. Let us stress that this result (for general k) is not obvious as strategic complementarity could a priori induce the existence of the entrenchment equilibrium even for arbitrarily large s/b. Checking that $\rho^{\mathrm{e}} < \infty$ thus requires some computations, in particular as the majority size has different transition probabilities from the perspective of a majority member and from the one of a minority member (due to a member's conditioning on still being a member in the next periods).

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

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

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

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 would exist $\overline{\rho} > 0$ such that for any $s/b > \overline{\rho}$, there exists a strictly profitable deviation from entrenchment to meritocracy, which would yield the

[61] The proof also yields that $\rho^{\mathrm{e}\dagger}|_{s^\dagger > b} < \infty$ (thus in particular for $x^\dagger \geq 1/2$), where ρ^{e} defined in Proposition 9 (see Section 4.2.1).

result: $\rho^{\mathrm{e}} < \infty$. We thus show that for $b = 0$:

$$s - \delta \frac{k-1}{N-1}(u^{\mathrm{e}}_{k-1} + u^{\mathrm{e}}_{k}) > 0 \tag{20}$$

Using (3)-(4) and (5)-(11), the above inequality can be written as

$$\frac{\delta x \frac{k-1}{N-1}}{1 - \delta x - \delta(1-x)\left[\frac{k+1}{N-1} + \frac{k-2}{N-1}a_{k+1}\right]} \tag{21}$$
$$\times\left(1 - \frac{\delta x \frac{k}{N-1}}{1 - \delta(1-x)\left[\frac{k}{N-1} + \frac{k-2}{N-1}b_{k+1}\right]} - \frac{\delta x \frac{k-2}{N-1}}{1 - \delta(1-x)\left[\frac{k+1}{N-1} + \frac{k-3}{N-1}b_{k+2}\right]}\right) < 1$$

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

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

and

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

Indeed, computations using (3)-(4) and (5)-(11) for the entrenchment equilibrium, give that:

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

and thus, by rearranging,[62] (20) is equivalent to (21).

We thus show that for any $x \in (0,1)$ and $\delta \in [0,(N-1)/N)$, inequality (21) is satisfied.[63]. By construction, $(a_{k+l})_{l=1}^{k-2}$ and $(b_{k+l})_{l=1}^{k-2}$ 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.[64] Therefore, the term on the first line (resp. second line) in (21) is strictly increasing (resp. decreasing) with respect to x and δ.

Using the inequality $b_{k+1} < b_{k+2} < 1$, a sufficient condition for (21) to be satisfied is

$$\delta x \frac{k-1}{N-1}\left(1 - \delta x \frac{N-2}{N-1} - \delta(1-x)\left[\frac{k}{N-1} + \frac{k-2}{N-1} b_{k+1}\right]\right) \tag{22}$$
$$/\left[\left(1 - \delta(1-x)\left[\frac{k}{N-1} + \frac{k-2}{N-1} b_{k+1}\right]\right)\left(1 - \delta x - \delta(1-x)\left[\frac{k+1}{N-1} + \frac{k-2}{N-1} a_{k+1}\right]\right)\right] < 1$$

or equivalently,

$$\delta x \frac{k-1}{N-1}\left(1 - \delta x \frac{N-2}{N-1} - \delta(1-x)\left[\frac{k}{N-1} + \frac{k-2}{N-1} b_{k+1}\right]\right) \tag{23}$$
$$-\left(1 - \delta(1-x)\left[\frac{k}{N-1} + \frac{k-2}{N-1} b_{k+1}\right]\right)\left(1 - \delta x - \delta(1-x)\left[\frac{k+1}{N-1} + \frac{k-2}{N-1} a_{k+1}\right]\right) < 0$$

The above inequalities are strictly stronger than (21) for any $x \in (0,1)$, and coincide with (21) in $x = 1$.

We now show that for any $x \in [0,1]$, (i) the LHS in (23) increases with δ over $[0,(N-1)/N]$, and (ii) its maximum (thus with $\delta = (N-1)/N$) is strictly negative.

[62]Using in particular that (3)-(4) imply:

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

[63]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 (20) is:

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

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

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

(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 $b_{k+1} < a_{k+1} < 1$, we have that $C_b < C_a < 1$. Using a downward induction argument on the sequences $(a_{k+l})_l$ and $(b_{k+l})_l$ yields that $\partial a_{k+1}/\partial\delta > \partial b_{k+1}/\partial\delta$.[65] As a consequence,

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

Denoting by $\varphi(\delta)$ the partial derivative of the LHS in (23) with respect to δ, we have after rearranging:

$$\begin{aligned}\varphi(\delta) = {}& x\Big(1+\frac{k-1}{N-1}\Big) + (1-x)(C_a+C_b) \\ & - 2\delta\Big[x(1-x)\Big(1+\frac{k-1}{N-1}\Big)C_b + (1-x)^2 C_a C_b + x^2\frac{k-1}{N-1}\frac{N-2}{N-1}\Big] \\ & + \delta(1-x)\frac{k-2}{N-1}\left(\frac{\partial a_{k+1}}{\partial \delta}\Big[1-\delta(1-x)C_b\Big] + \frac{\partial b_{k+1}}{\partial \delta}\Big[1-\delta(1-x)C_a - \delta x\Big(1+\frac{k-1}{N-1}\Big)\Big]\right)\end{aligned}$$

---

[65] 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 $l \in \{1, ..., k-3\}$,

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

Let $\psi(\delta) \equiv \varphi(\delta) - \delta(1-x)\dfrac{k-2}{N-1}\phi(\delta)$. We then note that $\psi(\delta) \geq 0$ [66], and therefore, $\varphi(\delta) > 0$ for any $x \in [0,1]$. Consequently, the LHS in (23) 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 (23) with $\delta = (N-1)/N$ is strictly negative. Indeed, the latter then writes as

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

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

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

Hence, a sufficient condition for the LHS in (23) to be strictly negative is that $1-2a_{k+1}+b_{k+1} > 0$. This actually holds,[67] which concludes the proof.

[66] Indeed, the expressions of ϕ and φ yield after rearranging:

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

where the last inequality stems from the fact that $k/(N-1) < C_b < C_a < 1$.

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

D Proof of Proposition 3

We first show the result for majority members. For any $i \in \{k, ..., N-1\}$, let $v_i \equiv V_i^{\mathrm{m}} - V_i^{\mathrm{e}}$. By construction, for any $i \geq k+1$, the recursive expressions of V_i^{m} and V_i^{e} yield:

$$\left[1 - \delta(1-x)\frac{i}{N-1} - \delta x\left(1 - \frac{i-1}{N-1}\right)\right]v_i = \delta(1-x)\left(1 - \frac{i}{N-1}\right)v_{i+1} + \delta x\frac{i-1}{N-1}v_{i-1}, \quad (24)$$

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 $\Delta \equiv x(s-b) + \delta x \frac{k-1}{N-1}(V_{k-1}^{\mathrm{m}} - V_{k+1}^{\mathrm{m}}) \geq 0$, this last inequality stemming from $s/b > \rho^{\mathrm{m}}$ (the inequality is strict whenever $x > 0$). Hence,

$$\left[1 - \delta\frac{k}{N-1}\right]v_k = \Delta + \delta\left(1 - \frac{k}{N-1}\right)v_{k+1} \quad (25)$$

Assume by contradiction that $v_{N-1} < 0$. Then, Equation (24) 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 (25) then yields $0 > (1-\delta)v_k > \Delta \geq 0$, which is a contradiction. Hence, $v_{N-1} \geq 0$, and by induction using Equation (24), $v_k \geq v_{k+1} \geq ... \geq v_{N-1} \geq 0$, which concludes the proof. Note that the inequalities are strict whenever $\Delta > 0$, i.e. whenever $x \in (0, 1/2)$ and $s \geq b$, or $x = 1/2$ and $s > b$. [Hence, whenever $x \in (0, 1/2)$, the result holds by continuity for $s < b$ with s/b in a neighborhood of 1. This observation (which holds analogously for the minority

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

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

value function) establishes the comparison mentioned in the text between the meritocratic and entrenchment selection rules.]

The result for minority members follows from analogous computations, noting that for $M \geq k+1$, meritocracy and entrenchment yield the same flow payoffs and transition probabilities, while in $M = k$,

$$V_{k-1}^{\mathrm{m}} - V_{k-1}^{\mathrm{e}} = x(s+b) + x\delta\left[\frac{k-2}{N-1}(V_{k-1}^{\mathrm{m}} - V_{k-2}^{\mathrm{m}}) + \frac{k}{N-1}(V_{k}^{\mathrm{m}} - V_{k-1}^{\mathrm{m}})\right]$$
$$+ \delta\left[\frac{k-2}{N-1}(V_{k-2}^{\mathrm{m}} - V_{k-2}^{\mathrm{e}}) + \frac{k+1}{N-1}(V_{k-1}^{\mathrm{m}} - V_{k-1}^{\mathrm{e}})\right],$$

where $V_{k-2}^{\mathrm{m}} \leq V_{k-1}^{\mathrm{m}} \leq V_{k}^{\mathrm{m}}$ by Lemma 2. Hence, $V_i^{\mathrm{m}} \geq V_i^{\mathrm{e}}$ for any $i \leq k-1$.

Lastly, as a by-product of the proof, we have that 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 $M = k$.[68]

E Proof of Lemma 3

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 $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$,

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

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

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 $(1-$

[68] The result for $i \leq k-1$ can be established using analogous computations to the case $i \geq k$, relying on the recursive expressions of the minority value functions.

$x)N\nu_k^{\mathrm{m}} = x(k+1)\nu_{k+1}^{\mathrm{m}}$. One has in particular that

$$\begin{cases} \nu_{k+1}^{\mathrm{m}}\Big[\frac{x}{1-x}\frac{k+1}{N} + 1 + \sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i \prod_{j=1}^{i}\frac{k-j}{k+1+j}\Big] = 1 \\ \nu_{k+1}^{\mathrm{e}}\Big[1 + \sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i \prod_{j=1}^{i}\frac{k-j}{k+1+j}\Big] = 1 \end{cases}$$

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

Remark. The ergodic probability for the majority size to be equal to k at the beginning of a period in the entrenched equilibrium writes as $\nu_{k+1}^{\mathrm{e}}(k+1)/N$, and thus by the above expression, decreases with k.

F Proof of Proposition 4

Let ρ^W be uniquely defined by

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

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

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

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

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

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

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

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

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[69]

$$N(N-1)\left[1 + \left(\frac{1-x}{x}\right)^{k-1} \prod_{j=1}^{k-1} \frac{k-j}{k+1+j}\right] > \frac{2}{1-x}\left[1 + k^2\left(\frac{1-x}{x}\right)^{k-1} \prod_{j=1}^{k-1} \frac{k-j}{k+1+j}\right].$$

G Asymmetric homophily benefits: Proof of Proposition 2'

We know from Lemma 1 that in any pure-strategy MPE, any majority plays either the canonical meritocracy strategy or the canonical entrenchment strategy.

Suppose $b_B < b_A < s$. The existence regions of canonical entrenchment and canonical meritocracy are a corollary of Proposition 2. So is the uniqueness of canonical entrenchment among MPEs in pure strategies whenever $s < \min\left(\rho^{\mathrm{m}} b_A, \rho^{\mathrm{e}} b_B\right)$.

Let us consider the case $\rho^{\mathrm{e}} b_B < s < \rho^{\mathrm{m}} b_A$. The canonical meritocratic strategy for type-B agents and the canonical entrenchment strategy for type-A agents constitute an MPE in pure strategies as $s < \rho^{\mathrm{e}} b_A$ and $s > \rho^{\mathrm{m}} b_B$ (since $\rho^{\mathrm{m}} < \rho^{\mathrm{e}}$). Uniqueness follows from Lemma 1.

Lastly, if $\rho^{\mathrm{m}} b_A < s < \rho^{\mathrm{e}} b_B$, then by Proposition 2, the meritocratic and entrenchment equilibria coexist as $\rho^{\mathrm{e}} > s/b_B > s/b_A > \rho^{\mathrm{m}}$. The same argument as in the proof of Proposition 3 yields the Pareto-comparison.

H Non-linear homophily benefits: Proof of Proposition 2"

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. (To alleviate the notation, as we consider nonlinear yet symmetric benefits, we omit the horizontal-group subscript $X \in \{A, B\}$.)

Let $\tilde{\mathcal{B}}(i)$ denote the per-period homophily benefit enjoyed by a member whose in-group

[69] 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

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

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

has size i (thus, in the linear case, $\tilde{\mathcal{B}}(i) \equiv (i-1)\tilde{b}$). In this Section – and only in this Section –, we change the definition of the value function: let now V_i be the forward-looking discounted sum of future homophily and quality payoff for a member with in-group size i *net of the quality stock alone* (and not of the homophily stock). Indeed, the current quality stock (sum of members' talent) is still irrelevant looking forward, and we thus take it out of the value function to alleviate the expressions. By contrast, the current homophily stock (majority size) affects the incremental lifetime homophily contribution of a new in-group member.

With this new definition, the (forward-looking net-of-quality-stock) continuation value function of a majority member at majority size M is given by

$$\tilde{\mathcal{B}}(M+1) + s_M + \delta\Big[\frac{M}{N-1}V_M + \Big(1 - \frac{M}{N-1}\Big)V_{M+1}\Big]$$

if the majority recruits the majority candidate with talent $s_M \in \{0, s\}$, and by

$$\tilde{\mathcal{B}}(M) + s_m + \delta\Big[\frac{M-1}{N-1}V_{M-1} + \Big(1 - \frac{M-1}{N-1}\Big)V_M\Big]$$

if the majority recruits the minority candidate with talent $s_m \in \{0, s\}$.

H.1 Concave homophily benefits

Let $\tilde{\mathcal{B}}(i)$ be strictly concave in the number of in-group members i, i.e. $\tilde{\mathcal{B}}(i+1) - \tilde{\mathcal{B}}(i)$ be strictly decreasing in i. Suppose $\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k) \leq \tilde{s}$. Hence,

$$0 \leq \tilde{\mathcal{B}}(N) - \tilde{\mathcal{B}}(N-1) < \tilde{\mathcal{B}}(N-1) - \tilde{\mathcal{B}}(N-2) < ... < \tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k) \leq \tilde{s}. \tag{26}$$

The proof follows from arguments similar to the ones in the proof of Lemma 1 (see Online Appendix A). Consider first a given value of V_{k-1} (in a well-chosen set), and the majority's optimal control problem given V_{k-1}. Let $(V_M(V_{k-1}))_{M\geq k}$ be the (unique) solution to this problem, i.e. such that for all $M \geq k$, the Bellman equation holds:

$$V_M = \mathbb{E}_{\tilde{v}_M, \tilde{w}_M}\Big[\max\Big\{\tilde{v}_M + \delta\Big(\frac{M}{N-1}V_M + \Big(1-\frac{M}{N-1}\Big)V_{M+1}\Big)\Big), \\ \tilde{w}_M + \delta\Big(\frac{M-1}{N-1}V_{M-1} + \Big(1 - \frac{M-1}{N-1}\Big)V_M\Big)\Big\}\Big],$$

where $\tilde{v}_M \in \{\tilde{\mathcal{B}}(M), \tilde{\mathcal{B}}(M)+s\}$ (resp. $\tilde{w}_M \in \{\tilde{\mathcal{B}}(M+1), \tilde{\mathcal{B}}(M+1)+s\}$) is the flow-homophily and flow-and-future-quality value to a majority member when a minority (resp. majority) candidate is recruited at majority size M.

Consider then $M = k$, and suppose that:

$$\delta \frac{k-1}{N-1}(V_{k+1}(V_{k-1}) - V_{k-1}) > s - (\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k)),$$

i.e. that, by the Bellman equation at $M = k$,

$$V_k(V_{k-1}) = \tilde{\mathcal{B}}(k+1) + \overline{x}s + \delta\left[\frac{k}{N-1}V_k(V_{k-1}) + \frac{k-1}{N-1}V_{k+1}(V_{k-1})\right].$$

Computations analogous to the ones in the proof of Lemma A.1 (see Online Appendix A) then yield that the value function from the canonical entrenchment strategy satisfies the above recursive equation for V_k and the Bellman equations for $M \geq k+1$. Indeed, for any $M \geq k+1$, the value function with the entrenchment strategies satisfies:

$$\begin{aligned} V^{\text{e}}_{M+1} - V^{\text{e}}_{M} =& (1-x)(\tilde{\mathcal{B}}(M+2) - \tilde{\mathcal{B}}(M+1)) + x(\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M)) \\ &+ \delta(1-x)\left[\frac{M}{N-1}(V^{\text{e}}_{M+1} - V^{\text{e}}_{M}) + \left(1 - \frac{M+1}{N-1}\right)(V^{\text{e}}_{M+2} - V^{\text{e}}_{M+1})\right] \\ &+ \delta x\left[\frac{M-1}{N-1}(V^{\text{e}}_{M} - V^{\text{e}}_{M-1}) + \left(1 - \frac{M}{N-1}\right)(V^{\text{e}}_{M+1} - V^{\text{e}}_{M})\right] \end{aligned}$$

while for $M = k$,

$$\begin{aligned} V^{\text{e}}_{k+1} - V^{\text{e}}_{k} =& (1-x)[\tilde{\mathcal{B}}(k+2) - \tilde{\mathcal{B}}(k+1)] + xs \\ &+ \delta(1-x)\left[\frac{k}{N-1}(V^{\text{e}}_{k+1} - V^{\text{e}}_{k}) + \frac{k-2}{N-1}(V^{\text{e}}_{k+2} - V^{\text{e}}_{k+1})\right]. \end{aligned}$$

Hence, using the same recursive technique as in Online Appendix A), $V^{\text{e}}_{M+1} \geq V^{\text{e}}_{M}$ for all $M \geq k$, which implies that the majority's recruitment of the in-group candidate whenever he is at least as talented as the out-group candidate is optimal at any $M \geq k$. Moreover, by the same type of argument, the sequence $(V^{\text{e}}_{M+1} - V^{\text{e}}_{M})_{M \geq k}$ is decreasing.[70] Hence, by the same argument as in the proof of Lemma 1 (Online Appendix A.3),

$$\begin{aligned} &\delta\left[\frac{k}{N-1}(V^{\text{e}}_{k+1} - V^{\text{e}}_{k}) + \frac{k-2}{N-1}(V^{\text{e}}_{k+2} - V^{\text{e}}_{k+1})\right] \\ &\qquad \leq \frac{\delta\frac{N-2}{N-1}}{1 - \delta(1-x)\frac{N-2}{N-1}}\Big((1-x)[\tilde{\mathcal{B}}(k+2) - \tilde{\mathcal{B}}(k+1)] + xs\Big). \end{aligned}$$

[70]The result can be shown by contradiction and by induction, proceeding as in the proof of Lemma 2 and using (26).

As a consequence, since by construction, $s = \tilde{s}/(1-\delta(N-2)/(N-1))$ and by (26), $\tilde{s} > \tilde{\mathcal{B}}(k+2) - \tilde{\mathcal{B}}(k+1)$, the above inequality yields that:

$$\delta\Big[\frac{k}{N-1}(V^{\mathrm{e}}_{k+1} - V^{\mathrm{e}}_{k}) + \frac{k-2}{N-1}(V^{\mathrm{e}}_{k+2} - V^{\mathrm{e}}_{k+1})\Big] \leq s - \big[\tilde{\mathcal{B}}(k+2) - \tilde{\mathcal{B}}(k+1)\big].$$

Since the sequences $(\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M))_{M\geq k}$ and $(V^{\mathrm{e}}_{M+1} - V^{\mathrm{e}}_{M})_{M\geq k}$ are decreasing, we have for all $M \geq k+1$,

$$s - \big(\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M)\big) \geq \delta\Big[\frac{M-1}{N-1}(V^{\mathrm{e}}_{M} - V^{\mathrm{e}}_{M-1}) + \Big(1 - \frac{M}{N-1}\Big)(V^{\mathrm{e}}_{M+1} - V^{\mathrm{e}}_{M})\Big],$$

and thus the sequence (V^{e}_{M}) satisfies the Bellman equations for $M \geq k+1$.

Therefore, $V_M(V_{k-1}) = V^{\mathrm{e}}_{M}$ for all $M \geq k$. It can then be checked that the only strategy consistent with this value function is the one of canonical entrenchment.

Similarly, if on the opposite, the solution to the Bellman equations satisfies:

$$\delta\frac{k-1}{N-1}(V_{k+1}(V_{k-1}) - V_{k-1}) < s - \big(\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k)\big), \tag{27}$$

the same arguments as the ones used above and in the proofs of Lemma 1 show that letting V^{m} denote the value function corresponding to the canonical meritocratic strategies, $V_M(V_{k-1}) = V^{\mathrm{m}}_{M}(V_{k-1})$ for all $M \geq k$. And again, the canonical meritocratic strategy is the only one consistent with this value function.

Lastly, in a canonical equilibrium, $V_{N-1} \geq ... \geq V_{k+1} \geq V_k \geq V_{k-1}$. Hence in particular, a necessary condition for the equilibrium to be canonical meritocratic, i.e. for inequality (27) to hold is:

$$\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k) \leq s.$$

If instead, $\tilde{\mathcal{B}}(k+1) - \tilde{\mathcal{B}}(k) > s$, then canonical meritocracy cannot be an equilibrium. Moreover, by the same logic, if $\tilde{\mathcal{B}}(k+2) - \tilde{\mathcal{B}}(k+1) > s$, canonical entrenchment cannot be an equilibrium (as recruiting an untalented majority candidate against a talented minority candidate at majority size $k+1$ then yields a strictly profitable deviation), and only super-entrenchment can be an equilibrium.[71]

[71] Using the same recursive method as in the proof of Lemma 2 (see Online Appendix B), it can be shown that the value function generated by the level-l super-entrenchment strategies increases with majority size for any $M \geq k+l$: $V_{k+l} \leq V_{k+l+1} \leq ... \leq V_{N-1}$.

H.2 Convex homophily benefits

Let us note that, by considering flow incremental payoffs, the result clearly holds for any δ sufficiently low. In particular, for any δ sufficiently low, the unique equilibrium features meritocratic recruitments below the threshold, and entrenched ones above.

Let us now consider the general case ($\delta \in (0,1)$). Let $\tilde{\mathcal{B}}(\cdot)$ be convex, and let $\underline{M} \geq k$ be such that $\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M) < \tilde{s}$ (resp. $> \tilde{s}$) for any $M < \underline{M}$ (resp. $M \geq \underline{M}$). Let us further assume that $\underline{M} \leq N-2$.

Consider a given value of V_{k-1} (in a well-chosen set), and let $(V_M(V_{k-1}))_{M \geq k}$ be the (unique) solution to the majority's optimal control problem given V_{k-1}, i.e. such that for all $M \geq k$, the Bellman equation holds:

$$V_M = \mathbb{E}_{v_M, w_M}\Big[\max\Big\{\tilde{v}_M + \delta\Big(\frac{M}{N-1}V_M + \Big(1 - \frac{M}{N-1}\Big)V_{M+1}\Big),\\ \tilde{w}_M + \delta\Big(\frac{M-1}{N-1}V_{M-1} + \Big(1 - \frac{M-1}{N-1}\Big)V_M\Big)\Big\}\Big],$$

where $\tilde{v}_M \in \{\tilde{\mathcal{B}}(M), \tilde{\mathcal{B}}(M) + s\}$ (resp. $\tilde{w}_M \in \{\tilde{\mathcal{B}}(M+1), \tilde{\mathcal{B}}(M+1) + s\}$).

Consider the strategy consisting in always recruiting the majority candidate at any majority size $M \geq \underline{M}$ (entrenched recruitments), denoting by $(V_M^*)_{M \geq \underline{M}}$ its induced value function. For any $M \geq \underline{M}$,

$$V_{M+1}^* - V_M^* = \tilde{\mathcal{B}}(M+2) - \tilde{\mathcal{B}}(M+1) + \delta\Big[\frac{M}{N-1}(V_{M+1}^* - V_M^*) + \frac{N-M-2}{N-1}(V_{M+2}^* - V_{M+1}^*)\Big]$$

By convexity,

$$\tilde{\mathcal{B}}(N) - \tilde{\mathcal{B}}(N-1) \geq ... \geq \tilde{\mathcal{B}}(\underline{M}+1) - \tilde{\mathcal{B}}(\underline{M}) > \tilde{s}, \tag{28}$$

and thus the sequence $(V_{M+1}^* - V_M^*)_{M \geq \underline{M}}$ is positive and increasing. Moreover,

$$V_{M+1}^* - V_M^* \geq \frac{\tilde{\mathcal{B}}(M+2) - \tilde{\mathcal{B}}(M+1)}{1 - \delta\frac{N-2}{N-1}} \geq \frac{\tilde{\mathcal{B}}(\underline{M}+1) - \tilde{\mathcal{B}}(\underline{M})}{1 - \delta\frac{N-2}{N-1}} > s,$$

for any $M \geq \underline{M}$. As a consequence, for any $M \geq \underline{M} + 1$,

$$s - (\tilde{\mathcal{B}}(M+1) - \tilde{\mathcal{B}}(M)) < \delta\Big[\frac{M-1}{N-1}(V_M^* - V_{M-1}^*) + \Big(1 - \frac{M}{N-1}\Big)(V_{M+1}^* - V_M^*)\Big].$$

Hence, if the solution to the Bellman equations satisfies

$$V_M = \tilde{B}(M+1) + \overline{x}s + \delta\Big[\frac{M}{N-1}V_M + \frac{N-M-1}{N-1}V_{M+1}\Big]$$

for some $M \geq \underline{M} - 1$, then the strategy of entrenched recruitments at majority sizes $M' \geq M$ solves the Bellman equations for $M' \geq M$, and thus the unique solution to the Bellman equations is such that the majority candidate is always recruited, regardless of his talent, at all majority sizes $M' \geq M$.

Suppose by contradiction that there exists no such majority size, i.e. the solution to the Bellman equations corresponds to meritocratic recruitments at all majority sizes $M \geq \underline{M} - 1$. Then, the same arguments as in the linear case apply to majority sizes $M < \underline{M} - 1$, yielding that the solution of the Bellman equations is given either by the canonical meritocratic or canonical entrenched strategies. Yet, we now argue that with these strategies, the majority has a profitable deviation whenever it has majority size $N - 1$ and its in-group candidate is less talented.

Indeed, consider the value function $(V_M)_M$ generated by one such strategy. Building on previous arguments, $V_{M+1} \geq V_M$ for all $M \geq k$. In addition, using their recursive expressions yields for any $M \geq \max(k+1, \underline{M})$,

$$\begin{aligned}
V_M &= \overline{x}s + (1-x)\tilde{\mathcal{B}}(M+1) + x\tilde{\mathcal{B}}(M) + xs \\
&\quad + \delta(1-x)\Big(\frac{M}{N-1}V_M^* + \frac{N-M-1}{N-1}V_{M+1}^*\Big) + \delta x\Big(\frac{M-1}{N-1}V_{M-1}^* + \frac{N-M}{N-1}V_M^*\Big) \\
&< \overline{x}s + (1-x)\tilde{\mathcal{B}}(N) + x\tilde{\mathcal{B}}(N-1) - (N-M-1)\tilde{s} + xs \\
&\quad + \delta(1-x)\Big(\frac{M}{N-1}V_M^* + \frac{N-M-1}{N-1}V_{M+1}^*\Big) + \delta x\Big(\frac{M-1}{N-1}V_{M-1}^* + \frac{N-M}{N-1}V_M^*\Big)
\end{aligned}$$

where the inequality follows from (28). Then, using the recursive expression of V_M and that $V_{N-1} \geq V_{N-2} \geq ... \geq V_k$ yields that

$$\begin{aligned}
&\Big[1 - \delta(1-x)\frac{N-2}{N-1} - \delta x\Big]V_{N-2} \\
&\qquad < \overline{x}s + (1-x)\tilde{\mathcal{B}}(N) + x\tilde{\mathcal{B}}(N-1) - \tilde{s} + xs + \delta\frac{(1-x)}{N-1}V_{N-1} \\
&\qquad < \overline{x}s + \tilde{\mathcal{B}}(N) - \tilde{s} + x\frac{\delta\frac{N-2}{N-1}}{1-\delta\frac{N-2}{N-1}}\tilde{s} + \delta\frac{(1-x)}{N-1}V_{N-1}.
\end{aligned}$$

In addition, as the majority can secure at least the entrenchment payoff at $M = N - 1$,

$$(1-\delta)V_{N-1} \geq \overline{x}s + \tilde{\mathcal{B}}(N).$$

As a consequence,

$$\begin{aligned}\tilde{\mathcal{B}}(N)-\tilde{\mathcal{B}}(N-1)+\delta\frac{N-2}{N-1}\left[V_{N-1}-V_{N-2}\right] &> \tilde{s}+\delta\frac{N-2}{N-1}\left[\frac{\frac{1-\delta(1-x)\frac{N-2}{N-1}}{1-\delta\frac{N-2}{N-1}}}{1-\delta(1-x)\frac{N-2}{N-1}-\delta x}\right]\tilde{s} \\ &> \tilde{s}+\delta\frac{N-2}{N-1}\left[\frac{1}{1-\delta\frac{N-2}{N-1}}\right]\tilde{s} \\ &= s,\end{aligned}$$

i.e. recruiting the in-group candidate against a more talented out-group candidate is a profitable deviation for the majority when it has size $N-1$.

I Proof of Proposition 5

We first show the validity of the remark in the text on a blind principal ($\lambda = 0$), before establishing Proposition 5.

Let us first argue that given the organization members' entrenchment strategy, there is no current-period benefit for the principal to intervene at any majority size. Indeed, there is no current-period 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 favor of the majority candidate, it 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.[72]

Similarly, there is no current-period benefit for the principal to intervene when the majority is tight ($M = k$). Indeed, since a tight entrenched majority always votes for its own candidate, its vote carries no information on the candidates' respective talents. Hence, the principal picks the (or "a" if there is a tie) most talented candidate with probability $1 - 2x + (1/2)(2x) = 1 - x$, which is the same probability of the entrenched majority choosing the most talented candidate. However, when the majority is tight, it takes the homophily-maximizing decision with probability 1, while the principal can only do so with probability 1/2 as it does not observe horizontal types.

Let us now consider the distribution of future majority sizes, to show that the principal has no future-periods benefits from an intervention in the current period. At any majority size $M \geq k$, by picking the minority candidate instead of the majority one, the principal sets the organization on a path on which the distribution of future majority sizes is stochastically

[72] 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.

dominated at any future time by the one on the no-intervention/original path (using the same argument as in the proof of Proposition 2, see the proof of Lemma C.2 in Online Appendix C.2.1). Hence, at any future time, the organization is more likely to be in the tight-majority state ($M = k$) following the principal's appointment of the minority candidate. Yet, the (expected) current-period welfare-increment (for incumbent members from the current-period recruit) is minimal at state $M = k$, equal to $\overline{x}s(N-1) + bk$, while it is equal to $(\overline{x} + x)s(N-1) + (1-x)bM + xb(N-1-M) > \overline{x}s(N-1) + bk$ at any majority size $M \geq k+1$.[73]

Hence, a blind principal cannot outperform the majority's decision.[74]

We now turn to the proof of Proposition 5.

I.1 Proof of claim (i)

Let $\lambda > 0$ be the probability that the principal learns the quality of the candidates. We look for equilibria in which the principal intervenes whenever informed that meritocracy is violated (and only then). We consider the organization members' strategy and we show that, given such an intervention policy for the principal:

(a) for s/b sufficiently close to 1, there exists a profitable deviation from canonical entrenchment in $k+1$ (the unique equilibrium when s/b is close to 1 and $\lambda = 0$) toward super-entrenchment at level 1. The argument then extends to any level of super-entrenchment.

(b) for s/b sufficiently close to 1, full entrenchment is an equilibrium.

(c) for any s/b sufficiently close to 1, full-entrenchment equilibrium is the unique symmetric MPE in pure strategies.

In the next Section, to prove claim (ii), we will show that for any s/b sufficiently close to 1 and for any λ in an intermediate range, if the majority is (canonically, super or fully) entrenched, it is optimal for the principal to intervene whenever it is informed that the current-period recruitment violates meritocracy.

(a). For $i \geq k$, let V_i be the majority value function in the canonical entrenchment equilibrium when the principal is informed with probability λ and intervenes whenever informed

[73] At any majority size $M \geq k+1$, the principal "mistakenly" picking an untalented majority candidate instead of a talented minority candidate yields a lower aggregate welfare as in equilibrium, the majority itself prefers recruiting the talented minority candidate instead of an untalented majority candidate.

[74] Even if the principal observed horizontal types (but remained talent-blind), a non-intervention equilibrium would still exist as the principal could not strictly improve on the entrenched majority's choices. (The above argument would go through as in particular, when the entrenched majority is tight, its recruitment choice does not reveal any information about the quality of candidates.)

that meritocracy is violated. Consider a deviation from canonical entrenchment to super-entrenchment in $k+1$, i.e. the majority voting for 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

$$\begin{aligned}\Delta &\equiv (1-\lambda)\Big[b - s + \delta\Big(\frac{k+1}{N-1}V_{k+1} + \frac{k-2}{N-1}V_{k+2}\Big) - \delta\Big(\frac{k}{N-1}V_k + \frac{k-1}{N-1}V_{k+1}\Big)\Big] \\ &= (1-\lambda)\Big[b - s + \delta\Big(\frac{k-2}{N-1}u_{k+1} + \frac{k}{N-1}u_k\Big)\Big]\end{aligned}$$

where $u_i \equiv V_{i+1} - V_i$. The sequence $(u_i)_{1 \le i \le N-2}$ satisfies Equation (5) for any $i \ge k+1$, and Equation (11) for any $i \le k-3$, while

$$\begin{cases}\Big[1 - \delta(1-x)\frac{k}{N-1} - \delta x\lambda\frac{k-1}{N-1}\Big]u_k = x(1-\lambda)(s-b) + \delta(1-x)\frac{k-2}{N-1}u_{k+1} + \delta x\lambda\frac{k-1}{N-1}u_{k-1} \\ \Big[1 - \delta(1-x\lambda)\Big]u_{k-1} = (1-2x\lambda)b + \delta(1-x\lambda)\Big[\frac{k-2}{N-1}u_{k-2} + \frac{k-1}{N-1}u_k\Big] \\ \Big[1 - \delta(1-x)\frac{k+1}{N-1} - \delta x\lambda\frac{k-2}{N-1}\Big]u_{k-2} = -x(1-\lambda)(s+b) + \delta(1-x)\frac{k-3}{N-1}u_{k-3} + \delta x\lambda\frac{k}{N-1}u_{k-1}\end{cases} \tag{29}$$

Summing up on all indices yields[75]

$$\Big[1 - \delta\frac{x}{N-1} - \delta(1-x)\Big](u_1 + u_{N-2}) + (1-\delta)\sum_{i=2}^{N-3} u_i = (1-2x)b > 0 \tag{30}$$

Fix $b > 0$. For any $s \ge b$, the same argument as the one used in the proof of Lemma 2 yields $u_k > u_{k+1} > ... > u_{N-2} > 0$.[76] The differential deviation payoff is thus strictly positive if and only if

$$\delta\Big(\frac{k-2}{N-1}u_{k+1} + \frac{k}{N-1}u_k\Big) > s - b \tag{31}$$

Consequently, for $s = b$, (31) is satisfied as it writes

$$\delta\Big(\frac{k-2}{N-1}u_{k+1} + \frac{k}{N-1}u_k\Big) > 0$$

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

[75] Assuming $k \ge 4$. The expression for $k \in \{2, 3\}$ writes differently on the LHS but has the same implication.

[76] Put succinctly, one supposes by contradiction that $u_{N-2} \le 0$ and reaches a contradiction showing by induction, using (5) together with the above system, that this implies $u_{k-1} \le 0$. Then, if $u_1 \le 0$, (11) implies $u_i \le 0$ for all i, which contradicts (30); whereas if $u_1 > 0$, (11) implies $u_{k-1} > 0$ and we reach again a contradiction. Hence, $u_{N-2} > 0$ and the same induction argument using (5) thus brings the result.

entrenchment to super-entrenchment.

The same argument can be adapted to show 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.

(b). We now show the existence of the full entrenchment equilibrium for s/b sufficiently close to 1. Let now V_i, u_i correspond to the full-entrenchment strategies. 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

$$\Delta \equiv (1-\lambda)\Big[s - b - \delta\frac{N-2}{N-1}u_{N-2}\Big]$$

Explicit computation with (3)-(4) yield:

$$u_{N-2} = \delta(1-x\lambda)\frac{N-2}{N-1}u_{N-2} + \delta x\lambda\Big[\frac{N-3}{N-1}u_{N-3} + \frac{1}{N-1}u_{N-2}\Big]$$

and more generally for any $M \geq k$,

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

while for any $i \leq k-2$,

$$u_i = \delta(1-x\lambda)\Big[\frac{i-1}{N-1}u_{i-1} + \Big(1-\frac{i}{N-1}\Big)u_i\Big] + \delta x\lambda\Big[\frac{i-1}{N-1}u_i + \Big(1-\frac{i+1}{N-1}\Big)u_{i+1}\Big]$$

with

$$\Big[1-\delta(1-x\lambda)\Big]u_{k-1} = (1-2x\lambda)b + \delta(1-x\lambda)\Big[\frac{k-1}{N-1}u_k + \frac{k-2}{N-1}u_{k-2}\Big]$$

Summing up over all indices yields

$$\Big[1-\delta\Big(1-x\lambda\frac{N-2}{N-1}\Big)\Big]u_{N-2} + \Big[1-\delta\Big(1-\frac{x\lambda}{N-1}\Big)\Big]u_1 + (1-\delta)\sum_{i=2}^{N-3}u_i = (1-2x\lambda)b > 0 \tag{32}$$

Fix $b > 0$ and let $s = b$. The usual argument implies that $u_{N-2} > 0$.[77] Hence, the differential deviation payoff when the majority has size $N-2$ writes for $s = b$ as

$$\Delta = -(1-\lambda)\delta\frac{N-2}{N-1}u_{N-2} < 0.$$

[77] 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 (32). Therefore, $u_{N-2} > 0$, and by induction again $u_k > u_{k+1} > ... > u_{N-2} > 0$.

By continuity, the inequality holds for s/b in a neighbourhood of 1.

Since $u_k > u_{k+1} > ... > u_{N-2} > 0$, the most profitable (one-shot) deviation from full-entrenchment is when the majority has size $N-1$ and a talented minority candidate faces an untalented majority candidate. As a consequence, the above necessary condition is also sufficient.

Hence, full entrenchment is an equilibrium for s/b in a neighbourhood of 1.

(c). Lastly, we show that for s/b in a neighbourhood of 1, full-entrenchment equilibrium is the unique (pure-strategy) symmetric MPE. To this end, we show that, for s/b in a neighbourhood of 1, any (pure-strategy) symmetric MPE is monotonic, in the sense that a stronger majority makes more meritocratic recruitments. Together with (a), this establishes the uniqueness of full entrenchment.

Let $s = b > 0$. We show that in any symmetric MPE, the differential value function $(u_M)_{M\geq k-1}$ is strictly positive and strictly decreases with M. Since the difference between the payoffs from a meritocratic, resp. an entrenched recruitment at majority size M whenever the minority candidate is strictly more talented than the majority one writes as

$$s - b - \delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1 - \frac{M}{N-1}\right)u_M\right],$$

the monotonicity of $(u_M)_M$ implies the monotonicity of the equilibrium. Moreover, if the strict monotonicity of $(u_M)_M$ obtains for $s = b$, then by continuity, it persists for s/b in a neighbourhood of 1, which implies that, for s/b in such a neighbourhood, any symmetric MPE is monotonic.

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 $M \leq N-2$,

$$\begin{cases} u_M = \delta(1-x\lambda)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\lambda\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ \qquad \text{in an equilibrium in which the majority is entrenched in } M, M+1, \\ u_M = \delta(1-x)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ \qquad \text{in an equilibrium in which the majority is meritocratic in } M, M+1, \\ u_M = \delta(1-x)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\lambda\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] \\ \qquad \text{in an equilibrium in which the majority is entrenched (resp. meritocratic) in } M (\text{resp. } M+1), \\ u_M = \delta(1-x\lambda)\left[\frac{M}{N-1}u_M + \left(1-\frac{M+1}{N-1}\right)u_{M+1}\right] + \delta x\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_{M+1}\right] \\ \qquad \text{in an equilibrium in which the majority is meritocratic (resp. entrenched) in } M (\text{resp. } M+1), \end{cases}$$

together with similar expressions for u_i when $i \leq k-2$.

Let us first show that $u_i > 0$ for all $i \in \{k-1, ..., N-2\}$. We proceed by induction. Suppose by contradiction that $u_{N-2} < 0$. Then, the above recursive expressions imply that $u_{k-1} < u_k < ... < u_{N-2} < 0$.[78] Therefore, the majority is meritocratic at all majority sizes $M \geq k$.[79] But then, Lemma 2 implies that $u_{k-1} > u_k > ... > u_{N-2}$, a contradiction.

Suppose now (again by contradiction) that $u_{N-2} = 0$. The above recursive expressions then imply that $u_{k-2} < u_{k-1} = u_k = ... = u_{N-2} = 0$, and thus that $Vk-2 > V_{k-1} = V_k = ... = V_{N-1}$. However, this implies that at all majority sizes $M \geq k$, the majority recruits the majority candidate whenever he is at least as talented as the minority candidate (and the majority is indifferent when he is strictly less talented than the minority candidate): since the majority recruits its own candidate at least a fraction $1-x$ of the time, and the minority candidate at most a fraction $x < 1-x$ of the time at all majority sizes, V_M must be strictly higher than V_{k-1} for all $M \geq k$, a contradiction.

Therefore, $u_{N-2} > 0$, and the above system then implies that $u_i > 0$ for all $i \in \{k-1, ..., N-2\}$ as was to be shown.

Let us now show that $uk-1 > u_k > ... > u_{N-2}$. Using that $u_{N_2} > 0$ and $u_{N-3} > 0$, the

[78] Indeed, the above recursive expressions imply that there exists $(a, b) \in \{(1-x\lambda, x\lambda),\ (1-x, x),\ (1-x, x\lambda),\ (1-x\lambda, x)\}$ such that

$$u_{N-3} = \frac{1 - \delta\frac{N-2}{N-1}a - \frac{\delta}{N-1}b}{\delta\frac{N-3}{N-1}b}u_{N-2}.$$

Hence, $u_{N-2} < 0$ implies $u_{N-3} < u_{N-2} < 0$. The result obtains by induction on the majority size.

[79] Indeed, as $s = b$, the differential payoff between recruiting a talented minority candidate instead of an untalented majority candidate is equal to

$$-\delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1-\frac{M}{N-1}\right)u_M\right] > 0.$$

above system evaluated at $M = N-2$ implies that, for any (pure) strategies, $u_{N-2} < u_{N-3}$. Proceeding recursively for $M \geq k$, $0 < u_{M+1} < u_M$ and $u_{M-1} > 0$ implies by the same argument that $u_M < u_{M-1}$. Therefore, the sequence $(u_M)_{M\geq k}$ strictly decreases with M.

Remark: Non-ergodic welfare comparison. Proposition 3 yields that, whenever meritocracy co-exists with entrenchment, the former is preferred by all members of the organization at any majority size. The result goes through in this setting.

Namely, we show that for any $l \geq 2$, whenever super-entrenchment at level $l-1$ and super-entrenchment at level l co-exist in equilibrium, the former is preferred by all (current) members of the organization at any majority size. The result for majority members relies on the same computations as in the proof of Proposition 3 (see Online Appendix D), using that since super-entrenchment at level $l-1$ is an equilibrium[80],

$$s - b + \delta\Big(\frac{k+l}{N-1}u^{\mathrm{e},l-1}_{k+l} + \frac{k-l-2}{N-1}u^{\mathrm{e},l-1}_{k+l+1}\Big) \geq 0$$

where $u^{\mathrm{e},l-1}_i = V^{\mathrm{e},l-1}_{i+1} - V^{\mathrm{e},l-1}_i$ with $V^{\mathrm{e},l-1}_i$ the value function of being in a group of size i in the super-entrenchment at level $l-1$ equilibrium. The result for minority members also relies on analogous computations to the ones in the proof of Proposition 3 (see Online Appendix D): using the recursive expressions of the value function for minority members in a similar fashion, we have that $V^{\mathrm{e},l-1}_i \geq V^{\mathrm{e},l}_i$ for any $i \leq k-1$ if

$$s + b + \delta\Big(\frac{k-l-1}{N-1}u^{\mathrm{e},l-1}_{k-l-1} + \frac{k+l-1}{N-1}u^{\mathrm{e},l-1}_{k-l}\Big) \geq 0 \qquad (33)$$

We thus show that this inequality holds, using the recursive expressions of $(u^{\mathrm{e},l-1}_i)_i$. We distinguish two cases.
(1) if $u^{\mathrm{e},l-1}_{k-l} \geq 0$, then $0 \geq u^{\mathrm{e},l-1}_1 \geq u^{\mathrm{e},l-1}_2 \geq ... \geq u^{\mathrm{e},l-1}_{k-l}$.[81] Hence, inequality (33) holds.
(2) if $u^{\mathrm{e},l-1}_{k-l} \leq 0$, then $u^{\mathrm{e},l-1}_{k-l} \leq u^{\mathrm{e},l-1}_{k-l-1}$ and $u^{\mathrm{e},l-1}_{k-l} \leq u^{\mathrm{e},l-1}_{k-l+1}$. Indeed,

- consider the first inequality and suppose by contradiction that $u^{\mathrm{e},l-1}_{k-l-1} < u^{\mathrm{e},l-1}_{k-l}$. By the usual (contradiction and induction) argument, this implies that $u_1 < ... < u^{\mathrm{e},l-1}_{k-l} \leq 0$. However, by summing the recursive expressions of $u^{\mathrm{e},l-1}_i$ for $i = 1, ...k-l-1$, and

[80] Indeed, this implies that in equilibrium, meritocratic recruitments are the majority's best response whenever it has size $k+l$, hence the inequality.
[81] This can be shown by the usual argument, supposing by contradiction that $u^{\mathrm{e},l-1}_1 < 0$, which implies by the recursive expressions of $(u^{\mathrm{e},l-1}_i)_i$, that $0 > u^{\mathrm{e},l-1}_1 > ... > u^{\mathrm{e},l-1}_{k-l}$, hence a contradiction. Therefore, $u^{\mathrm{e},l-1}_1 \geq 0$, and the recursive expressions of $(u^{\mathrm{e},l-1}_i)_i$ now imply that $0 \geq u^{\mathrm{e},l-1}_1 \geq u^{\mathrm{e},l-1}_2 \geq ... \geq u^{\mathrm{e},l-1}_{k-l}$.

rearranging, we get

$$\Big[1-\delta\frac{x}{N-1}-\delta(1-x)\Big]u_1^{\mathrm{e},l-1}+(1-\delta)\sum_{i=2}^{k-l-2}u_i^{\mathrm{e},l-1}+\Big[1-\delta\Big(1-(1-x)\frac{k-l-1}{N-1}\Big)\Big]u_{k-l-1}^{\mathrm{e},l-1}$$
$$=\delta x\frac{k+l-1}{N-1}u_{k-l}^{\mathrm{e},l-1}>\delta x\frac{k+l-1}{N-1}u_{k-l-1}^{\mathrm{e},l-1}$$

Therefore, as $u_1^{\mathrm{e},l-1}<u_{k-l-1}^{\mathrm{e},l-1}$, rearranging implies that

$$\Big[2-\delta\Big(1+\frac{k+l}{N-1}\Big)\Big]u_{k-l-1}^{\mathrm{e},l-1}+(1-\delta)\sum_{i=2}^{k-l-2}u_i^{\mathrm{e},l-1}>0,$$

which is a contradiction, as $u_1^{\mathrm{e},l-1}<\ldots<u_{k-l}^{\mathrm{e},l-1}\leq 0$. Consequently, $u_{k-l}^{\mathrm{e},l-1}\leq u_{k-l-1}^{\mathrm{e},l-1}$.

- consider the second inequality and suppose by contradiction that $u_{k-l}^{\mathrm{e},l-1}>u_{k-l+1}^{\mathrm{e},l-1}$. Using the recursive expression of $u_{k-l+1}^{\mathrm{e},l-1}$, this implies that $u_{k-l+2}^{\mathrm{e},l-1}<u_{k-l+1}^{\mathrm{e},l-1}<0$, and by induction that $0>u_{k-1}^{\mathrm{e},l-1}$. However, we know from the above computations that $u_i^{\mathrm{e},l-1}>0$ for any $i\geq k-1$, and thus in particular, $u_{k-1}^{\mathrm{e},l-1}>0$, which contradicts the above implication. Hence, $u_{k-l}^{\mathrm{e},l-1}\leq u_{k-l+1}^{\mathrm{e},l-1}$.

Therefore, if $u_{k-l}^{\mathrm{e},l-1}\leq 0$, then $u_{k-l}^{\mathrm{e},l-1}\leq u_{k-l-1}^{\mathrm{e},l-1}$ and $u_{k-l}^{\mathrm{e},l-1}\leq u_{k-l+1}^{\mathrm{e},l-1}$, and thus

$$s+b+\delta\Big(\frac{k-l-1}{N-1}u_{k-l-1}^{\mathrm{e},l-1}+\frac{k+l-1}{N-1}u_{k-l}^{\mathrm{e},l-1}\Big)\geq s+b+\delta\frac{N-2}{N-1}u_{k-l}^{\mathrm{e},l-1},$$

and using the recursive expression of $u_{k-l}^{\mathrm{e},l-1}$,[82]

$$\Big[1-\delta\big[1-x(1-\lambda)\big]\frac{N-2}{N-1}\Big]u_{k-l}^{\mathrm{e},l-1}\geq-(1-\lambda)x(s-b).$$

As a consequence,

$$s+b+\delta\Big(\frac{k-l-1}{N-1}u_{k-l-1}^{\mathrm{e},l-1}+\frac{k+l-1}{N-1}u_{k-l}^{\mathrm{e},l-1}\Big)\geq s+b-\frac{\delta x(1-\lambda)(N-2)}{N-1-\delta\big[1-x(1-\lambda)\big](N-2)}(s-b)$$
$$\geq s+b-\frac{k-1}{k+1}(s-b)>0$$

Hence, inequality (33) holds in both cases ($u_{k-l}^{\mathrm{e},l-1}\lesseqgtr 0$), as was to be shown.

[82]Namely,

$$u_{k-l}^{\mathrm{e},l-1}=-(1-\lambda)x(s-b)$$
$$+\delta(1-x)\Big[\frac{k-l-1}{N-1}u_{k-l-1}^{\mathrm{e},l-1}+\frac{k+l+1}{N-1}u_{k-l}^{\mathrm{e},l-1}\Big]+\delta x\lambda\Big[\frac{k-l}{N-1}u_{k-l}^{\mathrm{e},l-1}+\frac{k+l-2}{N-1}u_{k-l+1}^{\mathrm{e},l-1}\Big]$$

I.2 Proof of claim (ii)

Suppose the principal maximizes quality. With the above arguments, for any $\lambda > 0$, the principal's strategy "overruling whenever informed that meritocracy is violated" and the majority's full entrenchment is an equilibrium for s/b close to 1. But it may not be unique – e.g., if λ is close to 0, no overruling and canonical entrenchment is an equilibrium for s/b close to 1. We argue that for λ sufficiently close to 1, "overruling whenever informed that meritocracy is violated" and full entrenchment is the unique equilibrium (with our equilibrium concept).

Specifically, let us show that for any λ sufficiently close to 1, for any s/b sufficiently close to 1, if the majority is (canonically, super or fully) entrenched, it is optimal for the principal to intervene whenever it is informed that the current-period recruitment violates meritocracy.

Note first that 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 meritocracy prevails.

As noted in the text, for $\lambda = 1$ (perfectly informed principal), the principal can reproduce the equilibrium path of (canonical) meritocracy, which strictly dominates in terms of quality the equilibrium path of any level of entrenchment. Hence, by continuity, keeping members' strategies fixed, for λ sufficiently close to 1, it is optimal for the principal to intervene whenever informed. Moreover, by the same argument as in our initial remark about a blind principal, whenever the principal is not informed, it cannot outperform an entrenched majority's choice in terms of aggregate welfare. Indeed, it selects the (or "a" in case of a tie) most talented candidate with the same probability as the majority in the current period, while making a choice that is suboptimal in terms of homophily payoffs in the current period, and its intervention induces a distribution over future majority sizes that is dominated in terms of future quality and homophily payoffs by the non-intervention distribution. Therefore, for λ close to 1, given the members' strategy (canonical, super- or full entrenchment), it is optimal for the principal to intervene if and only if it is informed that the current-period recruitment violates meritocracy.

Consequently, by claim (i), for s/b close to 1 (such that in particular, canonical entrenchment is the unique equilibrium under *laissez-faire*) and λ close to 1, the unique equilibrium is for the principal to intervene if and only if it is informed that meritocracy is violated, and for the majority to be fully entrenched.

Let us now show that, for s/b close to 1 and λ in an intermediate range, the principal achieves a higher ergodic quality when it commits not to intervene. To provide an intuition, consider s/b close to 1 and λ close to 1 so that in the unique equilibrium, the organization is fully-entrenched. Since the principal is only informed with probability strictly below 1, it cannot compensate all the "un-meritocratic" recruitments made by the fully-entrenched majority. Hence, at any majority size $M \geq k+1$, i.e. at which the majority would have made meritocratic recruitments under *laissez-faire*, the principal would be better off in terms of flow welfare, if it could commit not to intervene. By contrast, whenever the majority is tight ($M = k$), entrenchment would have prevailed under *laissez-faire*, and so the principal's intervention improves the flow welfare.

To make things precise, let us consider s/b sufficiently close to 1 and λ sufficiently close to 1 such that the unique equilibrium is for the majority to fully entrench and for the principal to intervene if and only if informed that the current-period recruitment violates meritocracy. Ergodic aggregate quality is then strictly higher when the principal commits not to intervene if and only if

$$N(N-1)(1-\lambda)xs > N(N-1)\nu^{\mathrm{e}}_{k+1}\frac{k+1}{N}xs, \qquad \text{i.e.} \qquad \lambda < 1 - \nu^{\mathrm{e}}_{k+1}\frac{k+1}{N},$$

which yields the result. The range of values of λ for which the result holds is non-empty in particular whenever x is sufficiently small, as ν^{e}_{k+1} goes to 0 when x goes to 0. It is also non-empty whenever δ is sufficiently small, as it is then a strictly dominating strategy for the principal to intervene whenever informed that the current recruitment violates meritocracy (as $s > b$ and quality benefits accrue to all organization members, while homophily ones only to in-groups).

J Proof of Proposition 6

Consider an entrenched organization, i.e. by the equilibrium selection (by Proposition 3, meritocracy thus prevails whenever it exists as an equilibrium), suppose $s/b < \rho^{\mathrm{m}}$. Let $T \equiv \eta y$ denote equal the minimal expected bonus per member needed for the organization to move from entrenchment to meritocracy[83]. For the sake of exposition, we first assume that the principal does not value members' homophily benefits, and thus letting ξ be the cost of public funds[84], the principal's objective function writes as the ergodic welfare with per-period

[83] Namely,

$$\frac{s^{+}(\eta, y)}{b} = \rho^{\mathrm{m}}, \qquad \text{i.e.} \qquad \eta y = \left(\frac{b}{s}\rho^{\mathrm{m}} - 1\right)\tilde{s} > 0$$

[84] 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

welfare given by $W = qS - \xi T$ [85]. Note that such an objective constitutes an *upper* bound on the admissible cost of a policy as (ergodic aggregate) homophily payoffs decrease when the organization goes from entrenchment to meritocracy (see Section 2.2.2). From previous computations on ergodic welfare, the (ergodic) efficiency gain from disentrenchment writes as $S^{\mathrm{m}} - S^{\mathrm{e}} = N(N-1)\nu_{k+1}^{\mathrm{e}}\frac{k+1}{N}x\frac{\tilde{s}}{1-\delta} > 0$. Rewarding quality is thus optimal for the principal if and only if

$$\xi\eta y N^2(\overline{x} + x) \leq N(N-1)\nu_{k+1}^{\mathrm{e}}\frac{k+1}{N}x\tilde{s}$$

where $N[\overline{x} + x]$ is the average number of talented members in a meritocratic organization, and ν_{k+1}^{e} 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:[86]

$$\xi \leq \frac{(k+1)(N-1)}{N^2}\cdot\frac{x\nu_{k+1}^{\mathrm{e}}}{\overline{x}+x}\cdot\frac{\tilde{s}}{\eta y} = \frac{(k+1)(N-1)}{N^2}\cdot\frac{x\nu_{k+1}^{\mathrm{e}}}{\overline{x}+x}\cdot\frac{\frac{s}{b}}{\rho^{\mathrm{m}} - \frac{s}{b}}$$

Note that the RHS strictly increases with s/b and goes to $+\infty$ as s/b goes to ρ^{m}.[87] 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 $B^{\mathrm{m}} < B^{\mathrm{e}}$).

K Proof of Proposition 7

K.1 Proof of claim (i)

Whenever a representation threshold is implemented, we refer to the "existence region of (constrained) meritocracy" as the set of values of s/b for which there exists an equilibrium in which recruitments are meritocratic (i.e. a talented candidate is always recruited against a strictly less talented candidate) whenever the representation threshold R is not binding. Put

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.

[85] This objective may be interpreted as the limit of the main objective for $q, \xi \to \infty$.

[86] By Inequality (19), a lower bound on the RHS of the above equation is given by

$$\frac{(k+1)(N-1)}{N^2}\cdot\frac{x\nu_{k+1}^{\mathrm{e}}}{\overline{x}+x}\cdot\frac{\tilde{s}}{\eta y} \geq \frac{(k+1)(N-1)^2}{(k-1)N^2}\cdot\frac{x(1-2x)\nu_{k+1}^{\mathrm{e}}}{\overline{x}+x}\cdot\frac{(1-\delta)}{\delta}$$

[87] 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 $(k+1)(N-1)\nu_{k+1}^{\mathrm{e}}/N^2$ decreases with N, the comparative statics of ρ^{m} with respect to N are non-trivial. Nonetheless, for N large, the first two terms $(k+1)(N-1)\nu_{k+1}^{\mathrm{e}}/N^2$ 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.

differently, we thus focus on the region of values for s/b for which majority alternance can exist in equilibrium. The result is (almost) immediate[88] for a representation threshold of 1. We thus focus on $R \geq 2$.

Consider a representation threshold $R = k - l$ with $l \in \{1, ..., k-2\}$, and denote by $\tilde{V}$ the value function from recruiting the most talented candidate (and breaking ties in favor of the majority candidate) at all majority sizes at which the representation threshold R is not binding (omitting the superscript m), and let $\tilde{u}_i \equiv \tilde{V}_{i+1} - \tilde{V}_i$. We will first show that the sequence $(\tilde{u}_i)_{i \geq k-1}$ is such that $\tilde{u}_{k+l-1} < 0$, and such that it satisfies at least one of the following assertions: (A_1) it decreases with i, or (A_2) it is always strictly negative.[89] As in the baseline case, the monotonicity property (A_1) would imply that the most tempting deviation from meritocracy to entrenchment is when the majority has size k and the minority candidate is strictly more talented than the majority candidate, while (A_2) would imply that deviations from constrained meritocracy to entrenchment at any size $i \geq k$ are non-profitable as they yield a deviation payoff bounded above by

$$-(s-b) + \delta\left[\left(1 - \frac{i}{N-1}\right)\tilde{u}_i + \frac{i-1}{N-1}\tilde{u}_{i-1}\right] < 0$$

Lastly, that $\tilde{u}_{k+l-1}$ is negative suggests that there may be profitable deviations from meritocracy with ties broken in favor of the majority candidate to meritocracy with ties broken in favor of the minority candidate when s/b is high enough (more on this below).

We first suppose by contradiction that $\tilde{u}_{k+l-1} \geq 0$. The usual induction argument relying on (5) 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 2, the above recursive expression for $\tilde{u}_{k+l-1}$ with (12) and (5) over indices k to $k+l-2$, and rearranging, yields on the LHS a weighted sum of $\tilde{u}_{k-1}, ..., \tilde{u}_{k+l-1}$ which is strictly positive, while on the RHS:

$$-xs - (1-x)b + (1-2x)b + \delta(1-x)\frac{k-2}{N-1}\tilde{u}_{k-2} = -x(s+b) + \delta(1-x)\frac{k-2}{N-1}\tilde{u}_{k-2},$$

and so $\tilde{u}_{k-2} > 0$. Summing (11) at $k-2$ to the above sum, and rearranging, yields on the

[88] As will be clear shortly, 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).

[89] 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.

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 (5)-(11)-(12) 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, $\tilde{u}_{k+l-1} < 0$.

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 (5) brings by induction that[90]

$$\tilde{u}_{k+l-1} < \tilde{u}_{k+l-2} < ... < \tilde{u}_{i^-+1} < \tilde{u}_{i^-} < 0 < \tilde{u}_{i^--1} < \tilde{u}_{i^--2} < ... < \tilde{u}_{k-1},$$

which yields that (A_1) holds. If $i^- \leq k-1$, then (A_2) holds.

Consequently, to show that with affirmative action, the existence region of (constrained) meritocracy expands towards lower values of s/b, it is sufficient to consider deviations from meritocracy to entrenchment when the majority is tight and faces an untalented majority candidate and a talented minority candidate, and to show that the condition for non-profitability is looser for any s/b with affirmative action than in the baseline setting (without affirmative action).

Explicit computations yield[91]

$$\begin{cases} \tilde{u}_{k+l-1} = -xs - (1-x)b + \delta x\left[\dfrac{k-l}{N-1}\tilde{u}_{k+l-1} + \dfrac{k+l-2}{N-1}\tilde{u}_{k+l-2}\right] \\ \tilde{u}_{k-l-1} = xs - (1-x)b + \delta x\left[\dfrac{k-l-1}{N-1}\tilde{u}_{k-l-1} + \dfrac{k+l-1}{N-1}\tilde{u}_{k-l}\right] \end{cases} \tag{34}$$

[90]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 $i = k+l-2$.

[91]By definition of affirmative action with representation threshold R, in any equilibrium

$$\begin{cases} \tilde{V}_{k+l} = \overline{x}s + \delta\left[\dfrac{k+l-1}{N-1}\tilde{V}_{k+l-1} + \dfrac{k-l}{N-1}\tilde{V}_{k+l}\right] \\ \tilde{V}_{k-l-1} = \overline{x}s + \delta\left[\dfrac{k-l-1}{N-1}\tilde{V}_{k-l-1} + \dfrac{k+l}{N-1}\tilde{V}_{k-l}\right] \end{cases}$$

Hence, in the meritocratic equilibrium,

$$\begin{cases} \tilde{u}_{k+l-1} = -xs - (1-x)b + \delta x\left[\dfrac{k-l}{N-1}\tilde{u}_{k+l-1} + \dfrac{k+l-2}{N-1}\tilde{u}_{k+l-2}\right] \\ \tilde{u}_{k-l-1} = xs - (1-x)b + \delta x\left[\dfrac{k-l-1}{N-1}\tilde{u}_{k-l-1} + \dfrac{k+l-1}{N-1}\tilde{u}_{k-l}\right] \end{cases}$$

Thus using (5) at $k+l-1$ and (11) at $k-l-1$, together with the fact that $u_i \geq 0$ for all i in the baseline setting, one gets[92]

$$\begin{cases} \left[1-\delta x\dfrac{k-l}{N-1}\right](\tilde{u}_{k+l-1}-u_{k+l-1}) < -xs-(1-x)b+\delta x\dfrac{k+l-2}{N-1}(\tilde{u}_{k+l-2}-u_{k+l-2}) \\ \left[1-\delta x\dfrac{k-l-1}{N-1}\right](\tilde{u}_{k-l-1}-u_{k-l-1}) < xs-(1-x)b+\delta x\dfrac{k+l-2}{N-1}(\tilde{u}_{k-l}-u_{k-l}) \end{cases}$$

Therefore, using (5) at $k+l-2$ and (11) at $k-l$, one gets

$$\begin{cases} \left[1-\delta x\dfrac{k-l+1}{N-1}-\delta(1-x)\dfrac{k+l-2}{N-1}-\delta(1-x)\dfrac{k-l}{N-1}\dfrac{\delta x\frac{k+l-2}{N-1}}{1-\delta x\frac{k-l}{N-1}}\right](\tilde{u}_{k+l-2}-u_{k+l-2}) \\ \quad < \dfrac{\delta(1-x)\frac{k-l}{N-1}}{1-\delta x\frac{k-l}{N-1}}\big[-xs-(1-x)b\big]+\delta x\dfrac{k+l-3}{N-1}(\tilde{u}_{k+l-3}-u_{k+l-3}) \\ \left[1-\delta x\dfrac{k-l}{N-1}-\delta(1-x)\dfrac{k+l-1}{N-1}-\delta(1-x)\dfrac{k-l-1}{N-1}\dfrac{\delta x\frac{k+l-1}{N-1}}{1-\delta x\frac{k-l-1}{N-1}}\right](\tilde{u}_{k-l}-u_{k-l}) \\ \quad < \dfrac{\delta(1-x)\frac{k-l-1}{N-1}}{1-\delta x\frac{k-l-1}{N-1}}\big[xs-(1-x)b\big]+\delta x\dfrac{k+l-2}{N-1}(\tilde{u}_{k-l+1}-u_{k-l+1}) \end{cases}$$

We begin by noting that

$$\frac{k-l}{N-1}\left[1-\delta x\frac{k-l-1}{N-1}\right] > \frac{k-l-1}{N-1}\left[1-\delta x\frac{k-l}{N-1}\right],$$

[92] 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 2 for details). Similarly for the second equation.

and[93]

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

Hence, we have that[94]

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

By downward (resp. upward) induction on $(\tilde{u}_i-u_i)$ for $i\geq k$ (resp. for $i\leq k-2$), we get that

$$C_1(\tilde{u}_{k-1}-u_{k-1})<-C_2xs-C_3(1-x)b<0 \tag{35}$$

where C_1, C_2 and C_3 are strictly positive constants that depend on the parameters k, l and x. Let us detail the induction argument. Using (5)-(11), we obtain two sequences $(a_j)_{0\leq j\leq l-2}$

[93] To see this, we observe that: $(k-l)(k+l-2)=(k-l+1)(k+l-1)-(2k-1)$, and as a consequence, using the above inequality,

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

The inequality thus obtains using that $\delta(1-x)\frac{k-l}{N-1}<1-\delta x\frac{k-l}{N-1}$.

[94] Note that

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

and $(b_j)_{0\leq j\leq l-2}$ such that for any $j \leq l-2$,

$$\begin{cases} a_j\big(\tilde{u}_{k+j}-u_{k+j}\big) \\ \quad < -[xs+(1-x)b]\dfrac{\delta(1-x)\dfrac{k-l}{N-1}}{1-\delta x\dfrac{k-l}{N-1}}\displaystyle\prod_{n=j+1}^{l-2}\left(\frac{\delta(1-x)}{a_n}\frac{k-n-1}{N-1}\right)+\delta x\frac{k+j-1}{N-1}(\tilde{u}_{k+j-1}-u_{k+j-1}) \\ b_j\big(\tilde{u}_{k-j-2}-u_{k-j-2}\big) \\ \quad < [xs-(1-x)b]\dfrac{\delta(1-x)\dfrac{k-l-1}{N-1}}{1-\delta x\dfrac{k-l-1}{N-1}}\displaystyle\prod_{n=j+1}^{l-2}\left(\frac{\delta(1-x)}{b_n}\frac{k-n-2}{N-1}\right)+\delta x\frac{k+j}{N-1}(\tilde{u}_{k-j-1}-u_{k-j-1}) \end{cases}$$

where

$$\begin{cases} a_{j-1}=1-\delta x\dfrac{k-j}{N-1}-\delta(1-x)\dfrac{k+j-1}{N-1}-\delta(1-x)\dfrac{k-j-1}{N-1}\dfrac{\delta x\dfrac{k+j-1}{N-1}}{a_j} \\ b_{j-1}=1-\delta x\dfrac{k-j-1}{N-1}-\delta(1-x)\dfrac{k+j}{N-1}-\delta(1-x)\dfrac{k-j-2}{N-1}\dfrac{\delta x\dfrac{k+j}{N-1}}{b_j} \end{cases}$$

We first note that by induction[95]

$$\forall j\leq l-1,\qquad \frac{\delta(1-x)}{a_j}\frac{k-j-1}{N-1}<1,\qquad \text{and}\qquad \frac{\delta(1-x)}{b_j}\frac{k-j-2}{N-1}<1 \tag{36}$$

Hence, using (12), the coefficient C_1 in (35) is given by

$$1-\delta(1-x)-\frac{\delta(1-x)}{a_0}\frac{k-1}{N-1}\delta x\frac{k-1}{N-1}-\frac{\delta(1-x)}{b_0}\frac{k-2}{N-1}\delta x\frac{k}{N-1}>1-\delta>0$$

Using (12) further implies that the coefficient C_3 in (35) is strictly positive. We then show by downward induction on j that for any $j\leq l-1$,

$$\frac{1}{a_j}\frac{k-j-1}{N-1}>\frac{1}{b_j}\frac{k-j-2}{N-1},$$

which will yield that $C_2>0$. The initialization ($j=l-1$) derives from the observation in footnote 95 (the case $j=l-2$ has also been established above). As for the induction, i.e. to

[95] The initialization with $j=l-1$ stems from the observation that

$$\delta(1-x)\frac{k-l}{N-1}<1-\delta x\frac{k-l}{N-1},\qquad \text{and}\qquad \delta(1-x)\frac{k-l-1}{N-1}<1-\delta x\frac{k-l-1}{N-1}$$

Moreover,

$$\delta(1-x)\frac{k-l}{N-1}\left[1-\delta x\frac{k-l-1}{N-1}\right]>\delta(1-x)\frac{k-l-1}{N-1}\left[1-\delta x\frac{k-l}{N-1}\right],\qquad \text{i.e.}\ \frac{1}{a_{l-1}}\frac{k-l}{N-1}>\frac{1}{b_{l-1}}\frac{k-l-1}{N-1}$$

show that $a_{j-1}(k-j-1) < b_{j-1}(k-j)$, we note that for any $j \geq 0$, the induction hypothesis implies that[96]

$$\frac{k-j-1}{N-1}\frac{k+j-1}{N-1}\frac{1}{a_j}\frac{k-j-1}{N-1} > \frac{k-j}{N-1}\frac{k+j}{N-1}\frac{1}{b_j}\frac{k-j-2}{N-1} - \frac{1}{a_j}\frac{k-j-1}{N-1}\frac{1}{N-1}$$

and thus, using (36),

$$\delta x\delta(1-x)\frac{k-j-1}{N-1}\frac{k+j-1}{N-1}\frac{1}{a_j}\frac{k-j-1}{N-1} > \delta x\delta(1-x)\frac{k-j}{N-1}\frac{k+j}{N-1}\frac{1}{b_j}\frac{k-j-2}{N-1} - \frac{\delta x}{N-1}$$

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,

$$s - b - \delta\frac{k-1}{N-1}(u_{k-1} + u_k) > s - b - \delta\frac{k-1}{N-1}(\tilde{u}_{k-1} - u_k),$$

i.e., the non-profitability condition for a deviation from meritocracy to entrenchment is (strictly) looser with a representation threshold R than without.

Remark: For s/b sufficiently high, meritocracy with reverse favoritism is an equilibrium: the majority always picks the most talented candidate and breaks ties in favor of the minority candidate. Let $b = 0 < s$. We first note that in the unconstrained meritocratic equilibrium, this implies that $u_i = 0$ for any $i \in \{1, ..., N-2\}$. The above computations then apply, switching the weights $1-x$ and x (except for the flow payoffs of $\tilde{u}_{k+l-1}$ and $\tilde{u}_{k-l-1}$ which remain respectively given by $-xs$ and xs). Hence, $\tilde{u}_i < 0$ for any $i \geq k-1$. Consequently, the deviation differential payoff from reverse-favoritism meritocracy to standard-favoritism meritocracy at majority size M is given by

$$\delta\left(\frac{M-1}{N-1}\tilde{u}^{\mathrm{m}}_{M-1} + \frac{N-1-M}{N-1}\tilde{u}^{\mathrm{m}}_{M}\right) < 0,$$

which yields the result. By contrast, the same argument implies that meritocracy with standard favoritism is no longer an equilibrium for s/b sufficiently high.[97]

[96]Indeed, we have that $(k-j-1)(k+j-1) = (k-j)(k+j) - (2k-1)$, and

$$\frac{k-j}{N-1}\left[1-\delta(1-x)\frac{k+j}{N-1}\right] = \frac{k-j-1}{N-1}\left[1-\delta(1-x)\frac{k+j-1}{N-1}\right] + \frac{1-\delta(1-x)}{N-1}$$

[97]Considering $b = 0 < s$, and observing that in the unconstrained meritocratic equilibrium, $u_i = 0$ for any $i \in \{1, ..., N-2\}$ and using the above computations in order to get that $\tilde{u}_i < 0$.

Remark: Non-ergodic welfare comparison. The same computations as in the proof of Proposition 3 (see Online Appendix D) apply. Therefore, whenever meritocracy and entrenchment coexist, at any majority size the meritocratic equilibrium is preferred to the entrenched equilibrium by (current) majority members. Building on analogous computations, it can be shown that the same preference also holds in several cases for all (current) minority members. By mimicking the argument in Online Appendix D, we have that $\tilde{V}_i^{\mathrm{m}} \geq \tilde{V}_i^{\mathrm{e}}$ for any $i \leq k-1$ if

$$s + b + \delta\Big(\frac{k}{N-1}\tilde{u}^{\mathrm{m}}_{k-1} + \frac{k-2}{N-1}\tilde{u}^{\mathrm{m}}_{k-2}\Big) > 0, \tag{37}$$

This inequality holds in particular whenever δ is small. This dominance in terms of non-ergodic welfare motivates the welfare analysis of Proposition 7-(ii).

K.2 Proof of claim (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:

$$\begin{cases} S^{\mathrm{e}} = N(N-1)\Big[\dfrac{k+1}{N}\nu^{\mathrm{e}}_{k+1}\overline{x} + \Big(1 - \dfrac{k+1}{N}\nu^{\mathrm{e}}_{k+1}\Big)(\overline{x}+x)\Big]\tilde{s} \\ S^{\mathrm{m,AA}} = N(N-1)\Big[\dfrac{l}{N}\nu^{\mathrm{m,AA}}_{N-l}\overline{x} + \Big(1 - \dfrac{l}{N}\nu^{\mathrm{m,AA}}_{N-l}\Big)(\overline{x}+x)\Big]\tilde{s} \end{cases}$$

and thus:

$$S^{\mathrm{m,AA}} - S^{\mathrm{e}} = N(N-1)\Big[\frac{k+1}{N}\nu^{\mathrm{e}}_{k+1} - \frac{l}{N}\nu^{\mathrm{m,AA}}_{N-l}\Big]x\tilde{s}$$

Explicit computations (see Lemma 3 and its proof in Section E) yield:

$$\begin{cases} \nu^{\mathrm{e}}_{k+1}\Big[1 + \displaystyle\sum_{i=1}^{k-1}\Big(\frac{1-x}{x}\Big)^i \prod_{j=1}^{i}\frac{k-j}{k+1+j}\Big] = 1 \\ \nu^{\mathrm{m,AA}}_{N-l}\Big[1 + \displaystyle\sum_{i=1}^{k-l-1}\Big(\frac{x}{1-x}\Big)^i \prod_{j=1}^{i}\frac{N-l+1-j}{l+j} + \Big(\frac{x}{1-x}\Big)^{k-l}\frac{k+1}{N}\prod_{j=1}^{k-l-1}\frac{N-l+1-j}{l+j}\Big] = 1 \end{cases}$$

Consequently, $S^{\mathrm{m,AA}} - S^{\mathrm{e}}$ has same sign as

$$(k+1)\left[1+\sum_{i=1}^{k-l-1}\left(\frac{x}{1-x}\right)^i\prod_{j=1}^{i}\frac{N-l+1-j}{l+j}+\left(\frac{x}{1-x}\right)^{k-l}\frac{k+1}{N}\prod_{j=1}^{k-l-1}\frac{N-l+1-j}{l+j}\right]$$
$$-\,l\left[1+\sum_{i=1}^{k-1}\left(\frac{1-x}{x}\right)^i\prod_{j=1}^{i}\frac{k-j}{k+1+j}\right]$$

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

$$(k+1)\left[1+\sum_{i=1}^{k-l-1}\left(\frac{x_{\mathrm{AA}}(l)}{1-x_{\mathrm{AA}}(l)}\right)^i\prod_{j=1}^{i}\frac{N-l+1-j}{l+j}+\left(\frac{x_{\mathrm{AA}}(l)}{1-x_{\mathrm{AA}}(l)}\right)^{k-l}\frac{k+1}{N}\prod_{j=1}^{k-l-1}\frac{N-l+1-j}{l+j}\right]$$
$$=l\left[1+\sum_{i=1}^{k-1}\left(\frac{1-x_{\mathrm{AA}}(l)}{x_{\mathrm{AA}}(l)}\right)^i\prod_{j=1}^{i}\frac{k-j}{k+1+j}\right]$$

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.

L Proof of Proposition 8

We use a fixed-point argument to prove the existence of a class of equilibria characterized by a weakly decreasing decision rule $(\Delta_M)_M$ [98]. Let $\overline{u}$ be given by

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

Note that $(\mathbb{E}\big[(s+b)\mathbf{1}\{\hat{s}-s \le b\}\big]+\mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s}-s > b\}\big])$ 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, $u_M \in [0, \overline{u}]$ 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

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

size $M \in \{k, ..., N-1\}$, it favors a majority candidate with (discounted) talent s against a minority candidate with (discounted) talent $\hat{s}$ if and only if[99]

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

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 $D : K \longrightarrow \big[0, \min(b + \delta\overline{u}, \overline{s})\big]^k, u \longmapsto (D_M)_{M\in\{k,\dots,N-1\}}$, 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 $T : [0, +\infty]^k \times [b, \overline{s}]^k \longrightarrow [0, +\infty]^k, ((V_M)_M, (\Delta_M)_M) \longmapsto (T_M)_M$, where

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

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 $V_{k-1} \geq 0$. Given 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

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

for any $M \in \{k-1, ..., N-2\}$ (with the convention that $T_{k-1}(V(u), D(u)) \equiv V_{k-1}$). 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

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

member can guarantee.[100] Hence, we have by construction that for any $u \in K$ and any $i \in \{k-1, ..., N-2\}$, $\Upsilon_i(u) \leq \overline{u}$. 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 $u \in K$. Rearranging the above expression for $T_M(V(u), D(u))$ yields:

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

We thus distinguish two cases.

(A) If $D_M(u) < \overline{s}$ for all $M \geq k$, then[101]

$$\begin{aligned} T_M(V(u), D(u)) &= \mathbb{E}\big[s\mathbf{1}\{\hat{s} - s \leq D_M\}\big] + \mathbb{E}\big[\hat{s}\mathbf{1}\{\hat{s} - s > D_M\}\big] + \mathbb{P}(\hat{s} - s \leq D_M)D_M \\ &\quad + \delta\Big[\frac{M-1}{N-1}V_{M-1} + \Big(1 - \frac{M-1}{N-1}\Big)V_M\Big] \\ &= h(D_M) + \delta\Big[\frac{M-1}{N-1}V_{M-1} + \Big(1 - \frac{M-1}{N-1}\Big)V_M\Big] \end{aligned} \tag{38}$$

Consequently, if $D_M(u) < \overline{s}$ [102], 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

$$\begin{aligned} \Upsilon_M(u) &= h(D_{M+1}) - h(D_M) + \delta\Big[\frac{M-1}{N-1}u_{M-1} + \Big(1 - \frac{M}{N-1}\Big)u_M\Big] \\ &= h(D_{M+1}) + h_1(D_M) - b \end{aligned} \tag{39}$$

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 $[b, +\infty) \cap \text{Supp}(\hat{s} - s)$ [103]. This set notably includes the set

[100] 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 $X \geq b$.

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

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

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

of continuous joint symmetric distributions[104], 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.[105]

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 $M \geq k$,

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

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

[104]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 $\Delta \in (0, \overline{s})$,

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

and thus $h'(\Delta) \in (1/2, 1)$ since $\int_0^{\overline{s}} F(s)dF(s) = 1/2$.

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

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

a minority candidate with a strictly positive probability: $\Delta_M < \overline{s}$ for any $M \geq k+1$.[106]

Consequently, we only need to consider the case where $D_{k+1}(u) < D_k(u) = \overline{s} < \infty$ [107]. We first show that $\Upsilon_k(u) \in [\Upsilon_{k+1}(u), \overline{u}]$. By construction,

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

and thus, since $D_{k+1} < \overline{s}$ implies that $T_{k+1}(V, D)$ is given by (38),

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

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[108]

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

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

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

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

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

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

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

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

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

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

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 $h(\overline{s}) = \mathbb{E}[s] + \overline{s}$.

This second inequality is looser than the condition[109] 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 now show that any equilibrium characterized by a sequence of cut-offs $(\Delta_M)_{M \geq k}$ is such that (a) $\Delta_M > b$ for any $M \geq k$, and (b) the sequence $(\Delta_M)_M$ is strictly decreasing.

(a) We first argue that in any equilibrium, $\Delta_M > b$ for any $M \geq k$. We show this by downward induction. Suppose that $\Delta_{N-1} \leq b$. Then[110], this implies that $u_{N-2} \leq 0$, i.e. $V_{N-2} \geq V_{N-1}$. Hence the continuation payoff for a majority of size $N-1$ is bounded below by δV_{N-1}. By deviating from Δ_{N-1} to the value that maximizes the flow payoff, a majority with size $N-1$ gets a utility greater than

$$\max_X \left\{ \mathbb{E}[(s+b)\mathbf{1}\{\hat{s}-s \leq X\}] + \mathbb{E}[\hat{s}\mathbf{1}\{\hat{s}-s > X\}] \right\} + \delta V_{N-1}$$

Hence this would imply that

$$(1-\delta)V_{N-1} \geq \max_X \left\{ \mathbb{E}[(s+b)\mathbf{1}\{\hat{s}-s \leq X\}] + \mathbb{E}[\hat{s}\mathbf{1}\{\hat{s}-s > X\}] \right\}$$

which is a contradiction as the RHS is the highest attainable flow payoff (and $\delta > 0$). Therefore $V_{N-1} > V_{N-2}$, and thus $\Delta_{N-1} > b$. Suppose now that $V_{M'+1} > V_{M'}$ for any $M' \geq M$, and that $V_M \leq V_{M-1}$. Therefore, the continuation payoff for a majority of size M is bounded below by δV_M. Hence, by deviating from Δ_M to the value that maximizes the flow payoff, a majority with size M gets a utility greater than

$$\max_X \left\{ \mathbb{E}[(s+b)\mathbf{1}\{\hat{s}-s \leq X\}] + \mathbb{E}[\hat{s}\mathbf{1}\{\hat{s}-s > X\}] \right\} + \delta V_M,$$

which again leads to a contradiction. Consequently, $u_{M-1} > 0$, while $u_M > 0$ by the induction

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

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

[110] Using that by construction,

$$\Delta_{N-1} = b + \delta \frac{N-2}{N-1} u_{N-2}$$

hypothesis. Hence, since by construction we have that either $\Delta_M = \overline{s} > b$, or

$$\Delta_M = b + \delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1 - \frac{M}{N-1}\right)u_M\right], \tag{40}$$

this implies that $\Delta_M > b$. By induction, the inequality holds for any majority size $M \geq k$.

(b) We thus show that the sequence $(\Delta_M)_{M\geq k}$ is strictly decreasing. We first consider the case where for any $M \geq k$, $\Delta_M < \overline{s}$, and therefore (40) holds, and

$$u_M = h(\Delta_{M+1}) + \Delta_M - h(\Delta_M) - b \tag{41}$$

Suppose by contradiction that $\Delta_{N-1} \geq \Delta_{N-2}$. By the above equations,

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

where the inequality derives from the strict monotonicity of h_1. Hence $\Delta_{N-1} \leq b$, which contradicts the above result. Therefore $\Delta_{N-1} < \Delta_{N-2}$. We henceforth proceed by induction. Suppose $\Delta_{M'+1} < \Delta_{M'}$ for any $M' \geq M$, and suppose by contradiction that $\Delta_M \geq \Delta_{M-1}$. By (41), using the monotonicity of h_1, we have that

$$u_M < \Delta_M - b, \qquad \text{and} \qquad u_{M-1} \leq \Delta_M - b,$$

and therefore,

$$\Delta_M < \left(1 - \delta\frac{N-2}{N-1}\right)b + \delta\frac{N-2}{N-1}\Delta_M,$$

i.e. $\Delta_M < b$, which is a contradiction. Hence for any $M \geq k$, $\Delta_{M+1} < \Delta_M$, as was to be shown.

We now consider the case where there exists $M \geq k$ such that $\Delta_M = \overline{s}$. This implies that $\Delta_{M+1} < \overline{s}$ as otherwise the explicit expressions of V_M and V_{M+1} would give that

$$\delta\left[\frac{M-1}{N-1}u_{M-1} + \left(1 - \frac{M}{N-1}\right)u_M\right] = 0, \qquad \text{and thus} \qquad \Delta_M = b < \overline{s},$$

which is a contradiction. Hence suppose by contradiction that $\Delta_{N-1} = \overline{s}$, then $\Delta_{N-2} < \overline{s} =$

Δ_{N-1}. Yet the above computations[111] thus yield that $\Delta_{N-1} \leq b < \overline{s}$, which is a contradiction. Therefore, $\Delta_{N-1} < \overline{s}$, and as a consequence, the above computations yield that $\Delta_{N-2} > \Delta_{N-1}$. We again proceed by induction. Suppose $\Delta_{M'+1} < \Delta_{M'}$ for any $M' \geq M$. If $\Delta_M < \overline{s}$, the above computations apply, yielding that $\Delta_M < \Delta_{M-1}$. Hence, suppose by contradiction that $\Delta_M = \overline{s} \geq \Delta_{M-1}$. As noted above, this implies that $\Delta_{M-1} < \overline{s}$ and (41) holds in $M-1$, and thus $u_{M-1} \leq \Delta_M - b$. Moreover, using the explicit expressions of V_{M+1} and V_M,

$$\begin{aligned} u_M &= h(\Delta_{M+1}) - h(\Delta_M) + \delta\Big[\frac{M-1}{N-1}u_{M-1} + \Big(1 - \frac{M}{N-1}\Big)u_M\Big] \\ &< \delta\Big[\frac{M-1}{N-1}u_{M-1} + \Big(1 - \frac{M}{N-1}\Big)u_M\Big] \end{aligned}$$

where the inequality follows from the monotonicity of h. Therefore, $u_M < u_{M-1}$. As a consequence,

$$\begin{aligned} \Delta_M = \overline{s} &\leq b + \delta\Big[\frac{M-1}{N-1}u_{M-1} + \Big(1 - \frac{M}{N-1}\Big)u_M\Big] \\ &< b + \delta\frac{N-2}{N-1}u_{M-1} < \Big(1 - \delta\frac{N-2}{N-1}\Big)b + \delta\frac{N-2}{N-1}\Delta_M, \end{aligned}$$

i.e. $\Delta_M < b < \overline{s}$, which is a contradiction. Hence, for any $M \geq k$, $\Delta_{M+1} < \Delta_M$, as was to be shown.

We then turn to showing that equilibria can be ranked from more to less meritocratic. Consider the class of equilibria characterized by a decreasing decision rule $(\Delta_M)_{M\in\{k,\dots,N-1\}}$. We refer in the following to an equilibrium by its decision rule $\Delta \equiv (\Delta_M)_{M\in\{k,\dots,N-1\}}$. 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,

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

[111] Using that as $\Delta_{N-1} = \overline{s}$,

$$\Delta_{N-1} \leq b + \delta\frac{N-2}{N-1}u_{N-2} \leq \Big(1 - \delta\frac{N-2}{N-1}\Big)b + \delta\frac{N-2}{N-1}\Delta_{N-1}$$

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

$$\begin{aligned} & h_{2,M}(\Delta_M) - h_{2,M}(\Delta'_M) \\ & \quad = \delta \frac{M-1}{N-1}\Big[h_1(\Delta_{M-1}) - h_1(\Delta'_{M-1})\Big] + \delta\Big(1 - \frac{M}{N-1}\Big)\Big[h(\Delta_{M+1}) - h(\Delta'_{M+1})\Big] \end{aligned} \tag{42}$$

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

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

We note that $h_{2,M}$ is strictly increasing over $[b, \bar{s}]$ [112]. By monotonicity of h_1, we get for $M = k+1$ that

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

Suppose by contradiction that $\Delta_{k+1} \geq \Delta'_{k+1}$. Then by monotonicity, $\Delta_{k+2} \geq \Delta'_{k+2}$. By summing Equation (42) in $k+1$ and $k+2$ and rearranging, we get that

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

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 $[b, \bar{s}]$, 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 (42) over the indices $k+1, ..., M$ and rearranging,

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

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

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 (42) over all these indices and rearranging yields

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

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 (42) 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 $M \geq k$".

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

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

We first note that $\Delta_k < \Delta'_k$ implies that $\Delta_k < \overline{s}$. Hence, for any $M \geq k$,

$$
\Delta_M = b + \delta \Big[\frac{M-1}{N-1} u_{M-1} + \frac{N-M-1}{N-1} u_M\Big],
$$

and therefore, for any $M \geq k$,

$$\left[1-\delta\left(1-\frac{M-1}{N-1}\right)[1-\mathbb{P}(\hat{s}-s \leq \Delta'_M)] - \delta\frac{M}{N-1}\mathbb{P}(\hat{s}-s \leq \Delta'_M)\right)\Bigg](V_M - V'_M)$$
$$= \mathbb{E}[(\hat{s}-s-\Delta_M)\mathbf{1}\{\Delta_M < \hat{s}-s \leq \Delta'_M\}] + \delta\mathbb{P}(\hat{s}-s \leq \Delta'_M)\left(1-\frac{M}{N-1}\right)(V_{M+1} - V'_{M+1})$$
$$+ \delta(1-\mathbb{P}(\hat{s}-s \leq \Delta'_M))\frac{M-1}{N-1}(V_{M-1} - V'_{M-1}) \tag{43}$$

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 ($\Delta'_k = \overline{s}$).

If $\Delta'_k < \overline{s}$, we need to adapt the arguments in the proof of Lemma 2 and Proposition 3. Suppose by contradiction that $V_{N-1} \leq V'_{N-1}$. Then equation (43) implies that $V_{N-2} - V'_{N-2} \leq V_{N-1} - V'_{N-1} \leq 0$, and thus by induction that $V_{k-1} - V'_{k-1} \leq V_k - V'_k \leq V_{k+1} - V'_{k+1} \leq ... \leq V_{N-1} - V'_{N-1} \leq 0$. However, since $\Delta_k < \Delta'_k < \overline{s}$, we have that

$$b + \delta\frac{k-1}{N-1}(V_{k+1} - V_{k-1}) < b + \delta\frac{k-1}{N-1}(V'_{k+1} - V'_{k-1}),$$

and thus, $V_{k-1} - V'_{k-1} > V_{k+1} - V'_{k+1}$, which contradicts the above inequality. Hence, $V_{N-1} \geq V'_{N-1}$, and (43) implies by induction that $V_{k-1} - V'_{k-1} \geq V_k - V'_k \geq ... \geq V_{N-1} - V'_{N-1} \geq 0$. Therefore, a more meritocratic equilibrium is preferred at any majority size by all majority members to a less meritocratic equilibrium.

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

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

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

$$\begin{aligned}
&\left[1-\delta\left(1-\frac{i-1}{N-1}\right)\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})-\delta\frac{i}{N-1}[1-\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})]\right](V_i-V'_i)\\
&=\mathbb{E}\left[\left[\hat{s}-s+b+\delta\left(\frac{i-1}{N-1}u_{i-1}+\frac{N-1-i}{N-1}u_i\right)\right]\mathbf{1}\{\Delta_{N-1-i}<\hat{s}-s\leq\Delta'_{N-1-i}\}\right]\\
&\quad+\delta\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i})\frac{i-1}{N-1}(V_{i-1}-V'_{i-1})\\
&\quad+\delta(1-\mathbb{P}(\hat{s}-s\leq\Delta'_{N-1-i}))\left(1-\frac{i}{N-1}\right)(V_{i+1}-V'_{i+1})
\end{aligned}\tag{44}$$

Hence, for δ close to 0, the expectation term on the RHS of (44) is strictly positive. Suppose by contradiction that $V_1 \leq V'_1$. Then, by induction, equation (44) yields that $V_k - V'_k \leq ... \leq V_1 - V'_1 \leq 0$. However, we know from above that $V_{k-1} - V'_{k-1} \geq 0$, hence a contradiction.

Therefore, $V_1 > V'_1$. Working in a similar fashion – by contradiction and by induction using (44) – yields that, for δ small, $V_i > V'_i$ for all $i \in \{1, ..., k-2\}$.

M Homogamic evaluation capability: Proof of Proposition 9 and complements

Before stating the general result (for $s^\dagger \lesseqgtr b$), let us build the intuition for the case: $s^\dagger > b$. For this case 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.]

Intuitively, when $s^\dagger > b$, the model becomes similar to our baseline setup, yet with two key differences:

(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 $s^\dagger - b < s - b$.

When $s^\dagger > b$, 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. Consequently, the minority candidate may be selected even though the two candidates are equally talented.

Proposition 9 in the text and its implications for welfare follow from the general results in the next Proposition and its Corollary.

Proposition M.1. ***(Canonical equilibria with homogamic evaluation capability)***

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

(ii) If $s^\dagger > b$, there exist finite thresholds $\rho^{e\dagger}$ and $\rho^{m\dagger}$ such that[113]

- *The entrenchment equilibrium – in which the majority always chooses the majority's candidate for $M = k$, while for all $M \geq k+1$, the majority chooses the majority's candidate if talented, and chooses the minority's candidate (of unknown talent) otherwise – 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 for all $M \geq k$ – exists if and only if $s/b \geq \rho^{m\dagger}$.*

Corollary M.2. ***(Canonical equilibria with homogamic evaluation capability: Welfare)*** *(i) Whenever $s^\dagger \leq b$, by leading to full entrenchment, homogamic evaluation capability lowers ergodic aggregate welfare relative to perfect information. (ii) As with perfect information, the meritocratic equilibrium dominates the entrenchment equilibrium in terms of ergodic aggregate welfare for any s/b whenever $x^\dagger$ is below or close to $1/2$, or close to 1.*[114] *Furthermore, the meritocratic and entrenchment equilibria with homogamic evaluation capability yield a lower ergodic aggregate welfare than their perfect-information counterparts.*

M.1 Proof of Proposition M.1

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.[115]

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 2 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

[113] If $b < s^\dagger$ and $x^\dagger \leq 1/2$, then $\rho^{m\dagger} < \rho^{e\dagger}$. 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 $s/b \geq x^\dagger/x$.

[114] Whenever they coexist, the meritocratic equilibrium is (still) preferred to the entrenchment equilibrium by all members at any majority size.

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

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 9, 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 $\rho^{\mathrm{e}} < +\infty$ [116] 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

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

where the probabilities $\pi^{\mathrm{m}\dagger}_{i,j}(t)$ are taken *(a)* following the meritocratic equilibrium strategies described in Proposition 9, 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 9 is analogous to that of Proposition 2. 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 $x^\dagger \leq 1/2$,

[116] Cf. Section C.2.3.

whereas it is decreasing (and convex) if $x^\dagger > 1/2$. Moreover, in this latter case it may then be that $V_k^{\mathrm{e}\dagger} < V_{k-1}^{\mathrm{e}\dagger}$. Nonetheless, for $x^\dagger > 1/2$, the sequence $(V_M^{\mathrm{m}\dagger})_{M\geq k}$ being decreasing implies that its differences $(V_{M+1}^{\mathrm{m}\dagger} - V_M^{\mathrm{m}\dagger})$ 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 2 applies to both equilibria[117], 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.

A (one-shot) deviation in majority size k from the entrenched strategy (defined in Proposition 9), i.e. picking the minority candidate (of unknown talent) instead of the untalented majority candidate, yields a payoff equal to:[118]

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

where the probabilities $\pi_{i,j}^{\mathrm{e}\dagger}(t)$ (resp. $\hat{\pi}_{i,j}^{\mathrm{e}\dagger}(t)$) are taken *(a)* following the entrenched equilibrium strategies described in Proposition 9, 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

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

[117] 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

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

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

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:

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

i.e. by rearranging,

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

The result for the existence region of the meritocratic equilibrium follows. Lastly, the proof for $\rho^{\mathrm{e},\dagger} < +\infty$ is in Section C.2.3.

Note moreover that Lemma C.2 holds with the transition probabilities $\pi^{\mathrm{e}\dagger}$ and $\pi^{\mathrm{m}\dagger}$ [119], 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).[120]

M.2 Proof of Corollary M.2

The same argument as the one used in the proof of Proposition (3) yields that, whenever they co-exist, the meritocratic equilibrium is preferred to the entrenchment equilibrium by all members at any majority size.

We now consider ergodic per-period aggregate welfare. We first show that with homogamic evaluation capability, meritocracy dominates entrenchment. To this end, we show that the result of Proposition 4, proved in Online Appendix F, holds replacing x with $x^{\dagger} \in [0,1]$. Analogous computations to the ones in Online Appendix F show that meritocracy dominates

[119] Indeed, the proof holds for any $x \in [0,1]$ as the stochastic matrices P and $\hat{P}$ (introduced in the proof of Lemma C.2) remain stochastically monotone and stochastically comparable (with P stochastically dominating $\hat{P}$) for any $x \in [0,1]$.

[120] 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 only hold for $x^{\dagger}$ sufficiently below 1.

entrenchment if and only if

$$N(N-1)x\left[\frac{x^\dagger}{1-x^\dagger}\frac{k+1}{N}+1+\sum_{i=1}^{k-1}\left(\frac{1-x^\dagger}{x^\dagger}\right)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\right]q\tilde{s}$$
$$>\frac{2x^\dagger}{1-x^\dagger}\left[1+\sum_{i=1}^{k-1}(i+1)^2\left(\frac{1-x^\dagger}{x^\dagger}\right)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\right]\tilde{b} \tag{45}$$

where $q \geq 1$. By Proposition 9, a necessary condition for meritocracy and entrenchment to exist is $b < s^\dagger$, i.e. $xs > x^\dagger b$. Therefore, a sufficient condition for (45) to be satisfied is

$$N(N-1)\left[\frac{x^\dagger}{1-x^\dagger}\frac{k+1}{N}+1+\sum_{i=1}^{k-1}\left(\frac{1-x^\dagger}{x^\dagger}\right)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\right]$$
$$>\frac{2}{1-x^\dagger}\left[1+\sum_{i=1}^{k-1}(i+1)^2\left(\frac{1-x^\dagger}{x^\dagger}\right)^i\prod_{l=1}^{i}\frac{k-l}{k+1+l}\right]$$

By Online Appendix F, the above inequality holds for any $x^\dagger \in [0, 1/2]$, as well as for $x^\dagger$ greater than but close to 1/2. Moreover, it clearly holds for $x^\dagger$ close to 1. [Numerical simulations suggest it holds for any $x^\dagger[0, 1]$.]

We then turn to the ergodic aggregate welfare comparison of homogamic evaluation capability with respect to perfect information: we show that meritocracy and entrenchment with homogamic evaluation capability are dominated by their perfect-information counterparts. We proceed as in Section 2.2.2.

We first note that in both equilibria, the ergodic distribution of majority sizes with perfect information first-order stochastically dominates the one with homogamic evaluation capability. Using the notation introduced in Section 2.2.2, we denote by $\nu_i^{\mathrm{r}\dagger}$ the ergodic probability of state i at the end of a period in regime $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$, and show that for $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$, the probability distribution $\{\nu_i^{\mathrm{r}}\}$ first-order stochastically dominates $\{\nu_i^{\mathrm{r}\dagger}\}$. Indeed, for $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$, consider the stochastic matrices P^{r} and $P^{\mathrm{r}\dagger}$ associated with the probability distribution over (end-of-period) majority sizes in equilibrium r respectively with perfect information and homogamic evaluation capability, from an outsider's perspective[121]. By construction, both P^{r} and $P^{\mathrm{r}\dagger}$ are stochastically monotone, and the two are stochastically comparable, with $P^{\mathrm{r}}_{i.}$ stochastically dominating $P^{\mathrm{r}\dagger}_{i.}$ for any row index i as $x^\dagger \geq x$. Therefore, the ergodic distribution of majority sizes in equilibrium r with perfect information (first-order) stochastically dominates the one with homogamic evaluation capability.

As a consequence, since the aggregate homophily payoff at a given majority size strictly increases with the majority size, perfect information yields a higher ergodic aggregate homophily

[121]Namely, for any $i, j \in \{1, ..., k\}$, the matrix component P^{r}_{ij} (resp. $P^{\mathrm{r}\dagger}_{ij}$) is the probability (from an outsider's perspective) that the (end-of-period) majority size moves from $k+i-1$ to $k+j-1$ from one period to another in equilibrium $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$ with perfect information (resp. with homogamic evaluation capability).

payoff than homogamic evaluation capability in equilibrium $\mathrm{r} \in \{\mathrm{e}, \mathrm{m}\}$. Moreover, by Section 2.2.2, the difference in aggregate per-period expected quality between perfect information and homogamic evaluation capability writes as

$$S^{\mathrm{r}} - S^{\mathrm{r}\dagger} = \begin{cases} 0 & \text{if r = m,} \\ N(N-1)\left[\nu_{k+1}^{\mathrm{e}\dagger} - \nu_{k+1}^{\mathrm{e}}\right]\dfrac{k+1}{N}x\tilde{s} & \text{if r = e.} \end{cases}$$

Hence, since the probability distribution $\{\nu_i^{\mathrm{e}}\}$ first-order stochastically dominates $\{\nu_i^{\mathrm{e}\dagger}\}$, $S^{\mathrm{r}} - S^{\mathrm{r}\dagger} \geq 0$. Therefore, meritocracy and entrenchment with homogamic evaluation capability are dominated by their perfect-information counterparts in terms of ergodic per-period aggregate welfare.

In order to establish the welfare claim in (i), we show that (perfect-information) entrenchment dominates full-entrenchment. The aggregate ergodic quality in the full-entrenchment equilibrium writes as $S^{\mathrm{f}} = N(N-1)\overline{x}\tilde{s}$, and thus using the computations of Section 2.2.2, the difference between the ergodic efficiency of an entrenched and fully-entrenched organization is given by

$$S^{\mathrm{e}} - S^{\mathrm{f}} = N(N-1)\left[1 - \nu_{k+1}^{\mathrm{e}}\frac{k+1}{N}x\right]\tilde{s}$$

Similarly, the difference ergodic homophily benefits is given by

$$B^{\mathrm{e}} - B^{\mathrm{f}} = \sum_{i=k+1}^{N} \nu_i^{\mathrm{e}}\left[i(i-1) + (N-i)(N-i-1) - N(N-1)\right]$$

Building on Online Appendix E, explicit computations[122] then yield that $q(S^{\mathrm{e}} - S^{\mathrm{f}}) + B^{\mathrm{e}} - B^{\mathrm{f}} > 0$ for any $s > b$, hence the result.

The second part of the welfare claim in (ii) stems from the explicit expressions of ρ^{m} and

[122] With the explicit expressions for the ergodic probabilities ν_i^{e} derived in Online Appendix E, $q(S^{\mathrm{e}} - S^{\mathrm{f}}) + B^{\mathrm{e}} - B^{\mathrm{f}}$ has the same sign as

$$\left[N(N-1)\left(1 - \frac{k+1}{N}x\right)q\tilde{s} + 2(k+1)(1-k)\tilde{b}\right] \\ + \sum_{i=1}^{k-1}\left(\frac{1-x}{x}\right)^i \prod_{l=1}^{i}\frac{k-l}{k+1+l}\left[N(N-1)q\tilde{s} + 2(i+k+1)(i-k+1)\tilde{b}\right]$$

The result obtains by noting that for any $x \leq 1/2$,

$$N(N-1)\left(1 - \frac{k+1}{N}x\right) > 2(k+1)(1-k),$$

and that for any $i \in \{1, ..., k-1\}$, $2(i+k+1)(i-k+1) > 2(k+2)(2-k) > -N(N-1)$.

$\rho^{\mathrm{m}\dagger}$ which imply that for δ close to 0,

$$\rho^{\mathrm{m}} = 1 + (1-2x)\frac{k-1}{N-1}\delta + O(\delta^2), \qquad \text{and} \qquad \rho^{\mathrm{m}\dagger} = \frac{x^\dagger}{x}\left[1 + (1-2x^\dagger)\frac{k-1}{N-1}\right]\delta + O(\delta^2),$$

and thus $\rho^{\mathrm{m}} < \rho^{\mathrm{m}\dagger}$. The first part derives from the above results, namely that meritocracy and entrenchment with homogamic evaluation capability are dominated by their perfect-information counterparts, and that meritocracy dominates entrenchment with homogamic evaluation capability as well as with perfect information.

N Proof of Proposition 10

Let us define the strategy corresponding to "super-entrenchment to level l" for any group with size i such that $i \geq k$ or $\Lambda(N-1-i) > 0$, as the strategy that coincides with the previous level-l super-entrenchment strategy for the majority (group size $i \geq k$), and that consists in always voting for the in-group candidate for the minority whenever $\Lambda(M) > 0$, i.e. whenever the minority is pivotal with a strictly positive probability. Formally, generalizing $\sigma(i)$ to be the probability that a group with size $i \geq 1$ votes for the out-group candidate when the latter is more talented than the out-group candidate, super-entrenchment strategies are defined by:

(i) $\sigma(i) = 0$ for all $i \in \{N-k-l, ..., k+l\}$ and $\sigma(i) = 1$ for $i \geq k+l+1$,

(ii) at any group size $i \geq N-k-l$, each group votes for its in-group candidate whenever she is equally or more talented than the out-group candidate.

We denote by V_i the corresponding value function and u_i its first-difference.

Proof for existence. Let $s = b > 0$. The usual computations[123] (see proof of Lemma 2) yield that for any $i \geq k+l$ and for any $i \leq k-2-l$, $u_i = 0$. The usual argument then applies: using that for group sizes $i \in \{k, ..., k+l-1\}$,

$$\left[1 - \delta\Lambda(i)\left(1 - \frac{i}{N-1}\right) - \delta(1-\Lambda(i+1))\frac{i}{N-1}\right]u_i$$
$$= [\Lambda(i) - \Lambda(i+1)]b + \delta\Lambda(i)\frac{i-1}{N-1}u_{i-1} + \delta(1-\Lambda(i+1))\left(1 - \frac{i+1}{N-1}\right)u_{i+1},$$

[123]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 $\leq k-2-l$.

while for group sizes $i \in \{k-2-l, ..., k-2\}$,

$$\left[1 - \delta\Lambda(N-i-2)\frac{i}{N-1} - \delta(1-\Lambda(N-i-1))\left(1-\frac{i}{N-1}\right)\right]u_i$$
$$= [\Lambda(N-i-2) - \Lambda(N-i-1)]b + \delta\Lambda(N-i-2)\left(1-\frac{i+1}{N-1}\right)u_{i+1}$$
$$+ \delta(1-\Lambda(N-i-1))\frac{i-1}{N-1}u_{i-1},$$

and lastly for group size $k-1$:

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

one first shows that $u_i > 0$ for any $i \in \{k-1-l, ..., k+l-1\}$. The only non-trivial case for profitable deviations is thus when the out-group candidate is more talented than the in-group one. Therefore, since in such a case, for $s = b$, the one-shot deviation differential payoff is given by

$$-\delta(1-\Lambda(i))\left[\left(1-\frac{i}{N-1}\right)u_i + \frac{i-1}{N-1}u_{i-1}\right] < 0$$

at group size $i \in \{k, ..., N-1\}$, and by

$$-\delta\Lambda(N-1-i)\left[\left(1-\frac{i}{N-1}\right)u_i + \frac{i-1}{N-1}u_{i-1}\right] < 0$$

at group size $i \in \{N-k-l, ..., k-1\}$, super-entrenchment to level l is an equilibrium.

The result obtains by continuity for s/b in a neighbourhood of 1.

Proof for 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 close to 1, 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 will then show that the minority best-replies to this strategy by voting for the in-group candidate whenever it is pivotal with a strictly positive probability, i.e. at any majority size $M \leq k+l-1$.

We henceforth consider a candidate equilibrium within the class of symmetric MPEs such that a stronger majority makes more meritocratic recruitments. We begin by noting that when $s = b$, a group's flow payoff whenever it is pivotal does not depend on its making a meritocratic or entrenched recruitment (as the difference between the two is equal to $x(s-b) = 0$) and is strictly positive (proportional to $\overline{x}s+b$). 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, for $s=b$, in any symmetric MPE such that a stronger majority makes more meritocratic recruitments

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

Suppose by contradiction that the above LHS is weakly lower than 0, and thus that the majority votes meritocratically at size $k+l$. Suppose first that $u_{k+l} \leq 0$. By monotonicity within the equilibrium class, the majority votes meritocratically at any size $i \geq k+l$, and thus the recursive expression of u_i for $i \geq k+l$ is given by (5) and yields[124] that $u_{k+l-1} \leq u_{k+l} \leq ... \leq u_{N-2} \leq 0$. Then, summing up (and rearranging) the recursive expression of u_{k+l-1} and u_i for $i \geq k+l$ (and rearranging) yields on the LHS a (positively) weighted sum of u_i, $i \geq k+l-1$, which is thus (weakly) negative, and on the RHS the sum of the flow-payoff term in u_{k+l-1}, which is strictly positive (as noted above, since $\Lambda(k+l-1)>0=\Lambda(k+l)$), and of a term proportional to u_{k+l-2}. Therefore, $u_{k+l-2}<0$. We proceed by induction in order to show that $u_i<0$ for any $i \in \{k-1,...,k+l-2\}$. Let $M \in \{k,...,k+l-2\}$, and suppose $u_i \leq 0$ for any $i \geq M$. Summing and rearranging as above the recursive expressions of the differential value function u_i over indices $i \in \{M,...,N-2\}$ 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 either to $\Lambda(M)(1-2x)b$, $\Lambda(M)b$ or $[\Lambda(M)+\Lambda(M+1)2x]b$, and is thus strictly positive. Therefore, $u_{M-1}<0$.

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

Let us consider the minority's incentives. Suppose by contradiction that $u_{k-l-1} \leq 0$. Then,

[124]This can be seen by supposing by contradiction that $u_{N-2}>0$, and reaching a contradiction using (5). The result then obtains by downward induction, using again (5).

the recursive expression of u_i for $i \leq k-l-2$ is given by (11) and yields 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 with a strictly positive probability. 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 (11) now yield that $0 < u_1 < ... < u_{k-l-1}$. 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 $i \in \{k+1, ..., k+l-1\}$.

Back to the majority, 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 $i \geq k+l$ (namely (5) as we suppose that the majority votes meritocratically at size $k+l$), we have that $u_{k+l-1} > u_{k+l} > u_{k+l+1} > ... > u_{N-2} > 0$. Consequently, for $s = b$,

$$s - b + \delta\left[\frac{k+l-1}{N-1}u_{k+l-1} + \left(1 - \frac{k+l}{N-1}\right)u_{k+l}\right] > 0$$

and thus the majority is entrenched when it has size $k+l$.[125] By continuity with respect to s/b, this inequality holds for any s/b sufficiently close to 1, yielding the majority's entrenchment at size $k+l$.

Hence, for $s/b > 1$ sufficiently close to 1, any candidate equilibrium such that a larger

[125] Note that, as $s = b$, entrenchment at size $k+l$ implies that $u_{k+l} = u_{k+l+1} = ... = u_{N-2} = 0$ (as all flow-payoff terms in the recursive expressions of u_i for $i \geq k+l$ are thus nil).

majority makes more meritocratic recruitments is such that the majority makes entrenched recruitments at majority sizes $M \leq k + l$, and using the same arguments as in the proof of Proposition 2, meritocratic recruitments at majority sizes $M \geq k + l + 1$.[126]

The usual recursive arguments (considering first $s = b$ then using the value functions' continuity with respect to s/b) then yield that for s/b sufficiently close to 1, the minority uniquely best-replies to such strategies by being entrenched whenever it is pivotal with a strictly positive probability.

This establishes, for s/b sufficiently close to 1, the uniqueness of the level-l super-entrenchment equilibrium within the class of equilibria such that a stronger majority makes more meritocratic recruitments.

O Proof of Proposition 11

We first show that when candidates reapply, meritocratic strategies do not sustain an equilibrium for s/b in some interval $[1, \rho^{\text{m}} + \epsilon)$ 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 $[1, \rho^{\text{m}} + \epsilon)$ with $\epsilon > 0$: an equilibrium may be observationally equivalent to a meritocratic equilibrium by exhibiting the same recruitment path, without necessarily being meritocratic off the equilibrium path (more on this below).

Let us define the meritocratic equilibrium as the equilibrium in which the majority always recruits the best candidate available[127] 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 a bit above ρ^{m} (and for all $s/b \in [1, \rho^{\text{m}}]$). 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 $\rho^{\text{m}} = 1$), and $s/b = 1$. Suppose the majority has size k, and no talented majority candidate available[128] but an infinite

[126] In fact, the argument implies that for s/b sufficiently close to 1, in any symmetric, possibly non-monotonic MPE, the majority makes entrenched recruitments when it has size $k + l$, and thus by the same arguments as in the proof of Proposition 2, and makes meritocratic recruitments when it has size $M \geq k + l + 1$. The requirement that a stronger majority makes more meritocratic recruitments then yields that for s/b sufficiently close to 1, the majority must make entrenched recruitments at sizes $M \leq k + l - 1$ too.

[127] Namely the best candidate among current-period and stored candidates, breaking ties in favor of in-group candidates as before.

[128] Namely, it has no such candidate in store, and the current-period majority candidate is untalented.

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 $\tilde{b}/2$ [129], 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[130]. 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

[129]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 equal to $\tilde{s}$ after it has relinquished it to the other group.

[130]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:

$$s - b - x(s-b)\frac{\delta k/(N-1)}{1-\delta k/(N-1)} > (1-x)(s-b) > 0$$

with respect to the payoff when candidates cannot reapply. Therefore, since in the absence of storability, we have the equivalence between the profitability of one-shot and permanent deviations[131], 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[132]. 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 $s/b \in [1, \rho^{\mathrm{m}}]$). Hence, 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.

P Hierarchies and the glass ceiling

For simplicity, we look at the continuous-time version of our model. 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 $X dt$. 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.

As noted in the text, 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). 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.
- *Differential mingling effect:* for organizational reasons, senior members tend to hang

[131] Hence, when candidates cannot reapply, the above repeated deviation yields a zero differential payoff for $s/b = \rho^{\mathrm{m}}$.

[132] 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.

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 $x < 1/[1 + J\chi^J/\chi^S] < 2x$, 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:[133]

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

In this region, the glass ceiling index is invariant with how pyramidal the organization is (J)[134], decreases with the frequency of senior-level vacancies (χ^S) and increases with the flow

[133] 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).

[134] 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.

of talented candidates (X). Covering all parameter regions, the glass ceiling index is monotonic with χ^S/X.[135]

Proposition P.1. ***(Glass ceiling)*** *In the hierarchical organization's steady state, hiring at the junior level is meritocratic. By contrast, there exists a glass ceiling for minority juniors.*

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. Talented women may then quit the organization due to a discouragement 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.

Q Negative homophily

As claimed in the text (see footnote 3), the case $\tilde{b} < 0$, corresponding to *negative homophily*, can be accommodated in our model. Indeed, the set of possible flow payoffs in any period still 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 (end-of-period) 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. $\tilde{s}/\tilde{b} < -1$): there always exists an equilibrium in which the majority votes for the most talented candidate with a tie-breaking rule in favor of the minority candidate.

Let us provide a few more details on the second case ($\tilde{s}/\tilde{b} < -1$). Indeed, the same computations as in the proof of Lemma 2 (see Online Appendix B) yield that, letting $u_i \equiv V_{i+1} - V_i$, with V_i the value function with such strategies, $0 < u_1 < ... < u_{k-1}$ and $u_{k-1} > ... > u_{N-2} > 0$, with

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

[135]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 $\gamma = +\infty$.

and thus in particular,

$$\left[1 - 2\delta x \frac{N-2}{N-1}\right] u_{k-1} < -(1-2x)b.$$

As a consequence, deviations that yield a lower current-period flow payoff, together with a lower (in a first-order stochastic sense) distribution of next-period in-group sizes are strictly unprofitable. Moreover, as $0 < u_{N-2} < ... < u_{k-1}$, the deviation differential payoff for the majority from picking its in-group candidate instead of an at-least-as-talented out-group candidate (hence opting for a higher distribution of next-period in-group sizes at the expense of a lower current-period flow payoff) is maximal when both candidates have the same talent and the majority has size k. It then writes as

$$b + \delta \frac{k-1}{N-1}(u_{k-1} + u_k) < b + \delta \frac{N-2}{N-1} u_{k-1} < 0$$

using the above upper bound on u_{k-1}. Therefore, such a deviation is never profitable for the majority, and thus these strategies form an equilibrium.