ISSN No. 2454 - 1427

CDE
May 2026

Meritocracy in the Face of Group Inequality

Rajiv Sethi
Barnard College, Columbia University
and the Santa Fe Institute

Rohini Somanathan
Email: rohini@econdse.org
Department of Economics,
Delhi School of Economics

Working Paper No. 363

Centre for Development Economics
Delhi School of Economics
Delhi- 110007

Meritocracy in the Face of Group Inequality[*]

Rajiv Sethi[†] Rohini Somanathan[‡]

May 2026

Abstract

Meritocratic systems are commonly understood as those that assign tasks to individuals who can best perform them. But future performance is not known prior to assignment, and must be inferred from other traits. We consider a model in which performance depends on two attributes—ability and training—where ability is endowed and unobserved and training is acquired and observed. The potential to acquire training depends on ability and resource access, so ability affects performance through two channels: indirectly through training and directly through the performance function. The population consists of two groups, each with the same ability distribution, but with differential access to resources. We show that performance-maximizing allocations are not necessarily group-blind or monotonic in training, and can involve greater representation of groups with lower resource access than would arise if the only criteria for selection were past performance. If policies are constrained to be monotonic due to the possibility of strategic applicant behavior, these representation effects can be mitigated or amplified, depending on the fraction of the population that is selected.

JEL Codes: D82, I24

Keywords: merit, affirmative action, group inequality.

[*]We thank Maitreesh Ghatak, for his comments on earlier versions of this paper, and the Center for Advanced Study in the Behavioral Sciences at Stanford University for fellowship support. We have also benefited from discussions with Kehinde Ajayi, Nageeb Ali, V. Bhaskar, Patricio Fernandez, Gerald Jaynes, Navin Kartik, Michael Lachmann, Glenn Loury, François Maniquet, Debraj Ray, John Roemer, Abhiroop Sarkar, Thomas Scanlon, and Sahana Subramanian.

[†]Department of Economics, Barnard College, Columbia University and the Santa Fe Institute.

[‡]Department of Economics, Delhi School of Economics.

1 Introduction

In June 2023, the Supreme Court of the United States held that explicit consideration of race in college admissions was unconstitutional, arguing that it "demeans the dignity and worth of a person to be judged by ancestry instead of by his or her own merit and essential qualities."[1] This claim—that meritocratic selection is inconsistent with the use of ancestry and other markers of group identity—is widespread in popular discourse, and often treated as self-evident. Our goal in this paper is to interrogate this claim and clarify the meaning of merit in a society where groups have unequal access to resources.

We argue that the standard conception of meritocracy ignores important informational asymmetries that characterize practical selection problems. Since organizations admit candidates before tasks are performed, the best they can do is to maximize *expected* performance. When measures of past achievement are noisy signals of future performance, taking group identity into account can lead to better inferences.

We model the selection problem facing an organization that seeks to maximize expected performance, and show why it might use group markers in this process. Performance is assumed to be a function of *ability*, which is unobserved, and *training*, which can be observed. There are two social groups and applicants in each group have the same distribution of ability, but differ in their training distributions because these depend additionally on their access to resources such as those embodied in school and neighborhood environments. We show that group-blind selection is efficient only in special cases, and that the representation of resource disadvantaged groups depends on the combined effect of human capital production functions, the share of total applicants selected, and the asymmetry in resource access across groups.

A critical assumption of our model is that some determinants of performance are observed, while others must be inferred. Able candidates who have been deprived of the types of resources that shape measurable metrics of achievement might exhibit large gaps between their past training and their future performance. Mindful of this, selectors may want to skip over some training levels to reach them. When social groups have differential access to resources, this can lead to selection policies that are group-contingent, and non-monotonic within groups. However, non-monotonic polices cannot generally be sustained in equilibrium if candidates can strategically underinvest in training. We explicitly

[1] See *Students for Fair Admissions v. Harvard*, 600 U.S. 181 (2023), quoting *Rice v. Cayetano*, 528 U.S. 495 (2000).

consider this possibility, and show that it can *amplify* the effects that arise without strategic underinvestment. That is, the representation of a group with lower access to resources can be even greater when strategic effects are taken into account than when they are not.

Our work is related to an extensive literature on selection rules in the face of group inequality, as well as some recent theoretical work on learning about characteristics that individuals may wish to disguise. In the next section, we discuss some connections between this literature and our paper.[2] Section 3 lays out our basic model, and Section 4 presents our main results on optimal selection and representation in the absence of strategic underinvestment. In Section 5, we allow agents to strategically underinvest in training and show how this alters the set of candidates chosen in equilibrium. We conclude in Section 6 with a discussion of open questions for future research.

It is worth emphasizing that our goal here is to characterize rather than to endorse performance-maximizing assignment rules, since these are critical benchmarks in the larger debate on social equality. The paper is motivated by current disagreements related to the meaning of meritocracy in academic and popular discourse, and by the laws and institutions that govern the implementation of meritocratic principles in unequal societies. By illustrating that group-conscious selection does not necessarily imply group-specific goals, we show that many of the institutions governing selection across the world are not necessarily inefficient simply because they are sensitive to identity. Their fairness and efficiency lies in the details of the mechanisms they use.

2 Related Literature

The concept of merit has been extensively discussed and variously interpreted. Richard Arneson (2015) uses the term "formal equality of opportunity" to refer to selection of the most qualified for any given position. John Roemer refers to assignment based on talent as the non-discrimination principle, according to which "race or sex, as such, should not count for or against a person's eligibility for a position, when race or sex is an irrelevant attribute insofar as the performance of the duties of the position is concerned" (Roemer, 1998, p.1). We use this definition as a starting point and illustrate why group identity may be relevant for selection for informational rather than intrinsic reasons.

[2]For a broader review of the literature and its connection to policy debates, as well as a numerical example that illustrates some of the effects identified in the present paper, see Sethi and Somanathan (2023).

Steven Durlauf, Thomas Scanlon and Amartya Sen each provide theoretical and philosophical foundations for a context-dependent and forward-looking definition of merit. In evaluating the efficiency implications of alternative assignment rules, Durlauf argues for a "move from merit as reward to merit as effectiveness" and argues that race-conscious admissions can improve efficiency (Durlauf, 2008, p. 139).[3] Scanlon (2018) emphasizes that the notion of talent relevant for procedural fairness depends on the nature of the positions for which individuals are being selected. Sen (2000) argues that merit cannot be defined independently of social goals and that the term should be assigned to actions that promote these goals. The origins of these debates go back at least as far as Aristotle, who argued that "where flute players are similar with respect to the art, aggrandizement in flutes is not granted to those who are better born... it is to one who is preeminent in the work that preeminence in the instruments should be granted."[4]

In the area of optimal selection rules, there is a sizable literature on identity-contingent admissions and hiring under conditions of imperfect information about potential candidates. Early work focused on statistical discrimination (Phelps, 1972; Arrow, 1973; Coate and Loury, 1993; Aigner and Cain, 1977; Cornell and Welch, 1996). More recent research has examined optimal selection rules when diversity and merit are both valued (Chan and Eyster, 2003; Fryer et al., 2008; Fryer and Loury, 2013). These papers ask how traditional or "sighted" affirmative action compares with color-blind affirmative action, which refers to policies that are not explicitly group-contingent, but are nevertheless motivated by diversity goals. In this literature, individual merit is treated as synonymous with some observable qualification such as a test score. For instance, Chan and Eyster (2003, p. 860) assume that the "expected academic promise... of a candidate with test score t is simply t: the higher a candidate's score, the higher her quality." In this framework, meritocratic allocations necessitate the application of a common qualification threshold to all members of the population, regardless of identity.

Cestau et al. (2017) study selection decisions made by an (unnamed) school district. They distinguish between *profiling*, which is an attempt to use demographic information to meet performance goals unrelated to diversity, and *affirmative action*, which involves preferential treatment for a group beyond levels justified by profiling. The former serves the goal of maximized performance, while the latter increases diversity relative to

[3]Cavanagh (2002) discusses alternative notions of merit at some length and argues that scholars using the term often confuse a conception of merit related to moral desert with one that relies on the value of individual characteristics in relation a particular occupation.

[4]*Politics* 111.12, 1283a1-3. Translation by Lord (2013).

performance-maximizing levels. They find that the district engages in both profiling and affirmative action with respect to family income, and engages in profiling but not affirmative action with respect to race. We show that bans on the use of group membership in the process of selection eliminate both performance profiling and affirmative action. Contrary to the rhetoric surrounding such initiatives, such mandates potentially block the implementation of meritocratic allocations.

There is an emerging literature on information transmission in which senders can disguise or distort their own attributes in order to influence the decision of a receiver (Frankel and Kartik, 2019, 2022; Dessein et al., 2023; Ball, 2024). These papers are concerned with equilibrium inferences, decisions, and the quality of information transmission at a high level of generality. Viewed from the perspective of this work, our applicants are senders with three characteristics—ability, resources, and group membership. Training corresponds to a message about ability, but resources constrain the space of messages that can be sent. Groups have different resource distributions and so group membership can affect the interpretation of messages.

A number of popular writings have highlighted the extent to which test scores reflect household and school resources rather than aptitude (Sandel, 2020; Markovits, 2019). Also related are studies by social psychologists that document the extent to which grades and test scores underestimate the potential of disadvantaged populations to perform. They argue that resources available to students in academic settings are both material and psychological. Building supportive psychological environments, a process that has been termed "affirmative meritocracy", can bridge the gap between measured academic grades and potential performance (Aronson and Steele, 2005; Walton et al., 2013). These writings support our assumption that the observable correlates of productivity are noisy measures of actual productivity.

To summarize, we define a meritocratic allocation as one that selects the most capable individuals from a pool of candidates. Like much of the literature on optimal selection, we assume performance is based on individual attributes. We depart from this literature in assuming that only some of these attributes can be observed and explore the role of group identity in maximizing expected performance.

3 The Model

Consider a population composed of two groups, 1 and 2, with shares s_1 and s_2 respectively. Individuals within each group differ in their ability a, and their past access to resources r. Suppose that both groups have the same distribution of ability, which is continuous with density $f(a)$ and distribution function $F(a)$. We assume that this distribution has support $[0,1]$ and that f is strictly positive in the interior of this domain.

The distribution of resources varies by group. Our main results are based on two resource levels, r_l and r_h where $r_h > r_l > 0$. The proportion of individuals in group i with access to the higher resource level is denoted by q_i. We assume $q_1 < q_2$ and refer to the first group as *disadvantaged*.[5]

Ability and resources are independently distributed and neither can be directly observed. We observe only group identity and a signal $t \geq 0$, which we refer to as *training*. This could be thought of as a measure of educational attainment, a test score, or any combination of productivity-relevant metrics. The highest attainable level of training for an individual is given by the continuous function $\tau(a, r)$, which is increasing in both arguments. That is, at each level of ability and resources, the level of training satisfies

$$t \leq \tau(a, r). \tag{1}$$

We are interested in the problem of selecting a fraction k of the applicant pool into scarce positions, and refer to k as *elite capacity*. Performance in these positions is increasing in both ability and training and is given by the function

$$p = \phi(a, t). \tag{2}$$

Ability therefore has both a direct and an indirect effect on performance, capturing the idea that past scores might inadequately represent the capacity of talented individuals to perform.

If performance depended only on ability, we would be in a pure signaling environment as in Spence (1973). However, when there are unobserved differences in access to resources across individuals and groups, and when both endowed ability and acquired training affect performance, some novel insights emerge.

[5]The assumption of two resource levels simplifies our analysis and exposition. In the appendix we consider an example with a continuous distribution of resources.

Any selection policy can be described by a pair of functions $\pi_i(t)$ that denote the probability of being selected conditional on exhibiting training t and belonging to group $i \in \{1, 2\}$. The selection policy that maximizes expected performance depends on the training and performance functions, $\tau(a, r)$ and $\phi(a, t)$, and on whether or not candidates choose training strategically. We refer to a selection policy as *monotonic* if, for each group i, there is a threshold training level which separates selected and rejected candidates.[6] A policy is *group-blind* if the likelihood of selection is independent of group membership, so $\pi_1(t) = \pi_2(t)$ at all levels of t. If a policy is not group-blind, we say that it is *group-sighted*. A group-sighted policy *favors the disadvantaged group* if $\pi_1(t) \geq \pi_2(t)$ at all t, with strict inequality at some t. In the next section we show that policies that maximize expected performance might be neither monotonic nor group-blind, and will favor the disadvantaged group under certain conditions.

Since the performance function is increasing in training, it is worth considering how a performance-maximizing selection policy could be non-monotonic in training. The answer lies in the inference about ability drawn at any given training level. If training is resource-intensive, while performance relies heavily on ability, an optimal selection policy might ignore some candidates with high training levels in favor of others with somewhat lower training. A non-monotonic selection policy of this kind, however, can create incentives for underinvestment in training. Those with low ability and high resources may underinvest in training to pool with those below them to increase their likelihood of selection. This in turn affects the equilibrium assignment rule. We explore strategic underinvestment in Section 5, but first characterize optimal selection in the absence of strategic training choices.

4 Main Results

4.1 Optimal Selection

In the absence of strategic underinvestment, the level of training for each applicant is uniquely determined as $t = \tau(a, r)$, and optimal selection rules are deterministic—each

[6]Selection at the threshold might be random if the distribution of training is discrete, which can happen if both ability and resources have discrete distributions, or if training is chosen strategically. See Section 5 for analysis of strategic training choices.

candidate is selected with probability zero or one.[7] For each t, we define the ability levels $\alpha_l(t)$ and $\alpha_h(t)$ that allow an applicant to achieve training t with low and high resources respectively:

$$t = \tau(\alpha_l(t), r_l) = \tau(\alpha_h(t), r_h).$$

Clearly $\alpha_l(t) \geq \alpha_h(t)$ with strict inequality when $t > 0$. Both α_h and α_l are strictly increasing in t; conditional on resource access, higher training reflects higher ability.

Meritocratic selection requires choosing applicants with the highest expected performance. In each group, those with $a = 1$ and high resources achieve the highest possible level of training, denoted $\bar{t} = \tau(1, r_h)$. Those with $a = 1$ and low resources achieve $t^* = \tau(1, r_l) < \bar{t}$. Therefore, those with training $t \in (t^*, \bar{t}]$ must have high resources and their performance is given by $\phi(\alpha_h(t), t)$.

At training levels $t < t^*$, there is a mix of applicants with low and high resources. Let $\gamma_i(t)$ denote the likelihood that an individual in group i, with training t, has high resources. The expected productivity of an applicant in group i with training $t < t^*$ is then

$$E(p_i|t) = \gamma_i(t)\phi(\alpha_h(t), t) + (1 - \gamma_i(t))\phi(\alpha_l(t), t). \tag{3}$$

The weight $\gamma_i(t)$ depends on the high resource share q_i, the ability distribution f, and the implicitly defined functions α_h and α_l.[8]

We next define sets of training levels that play an important role in our analysis. For each group i, define $\hat{p}_i$ as the highest level of expected performance for applicants with training in the range $[0, t^*]$:

$$\hat{p}_i \equiv \max_{t \leq t^*} E(p_i|t).$$

Note that $\hat{p}_i < E(p_i|\bar{t}) = \phi(\alpha_h(\bar{t}_i), \bar{t}_i)$, since the latter performance is delivered by the applicants with the highest feasible levels of *both* ability and training. Since performance is strictly increasing in training for $t > t^*$, there is a unique $\hat{t}_i \in [t^*, \bar{t})$ such that

$$\phi(\alpha_h(\hat{t}_i), \hat{t}_i) = \hat{p}_i. \tag{4}$$

If $\hat{p}_i > \phi(\alpha_h(t_i^*), t_i^*)$ then $\hat{t}_i > t^*$, otherwise $\hat{t}_i = t^*$. The following example illustrates both these possibilities.

[7] Optimal selection is non-deterministic in Chan and Eyster (2003) and Fryer and Loury (2013) despite continuous (and exogenously given) score distributions, because the authors assume that only monotonic policies are feasible. Relaxing this constraint results in generically deterministic selection (Ray and Sethi, 2010). When we consider strategic behavior in Section 5, it becomes necessary to allow for stochastic selection rules.

[8] A closed form expression for $\gamma_i(t)$ is derived in the appendix.

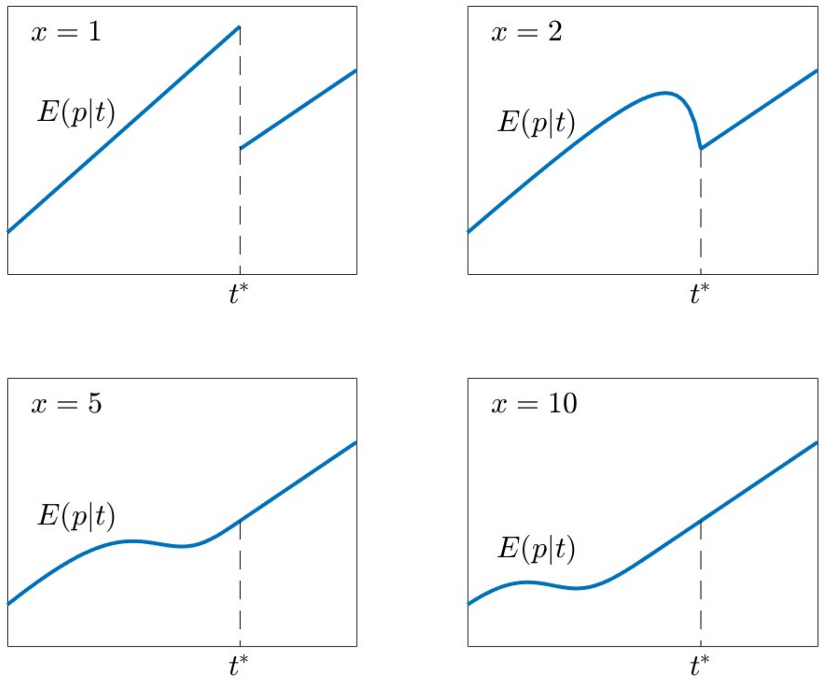

Figure 1: Expected performance conditional on training for ability distributions $a \sim \text{Beta}(x, x)$.

Example 1. *Suppose that the training and performance functions are $t = ar$, and $p = \beta a + (1 - \beta)t$ respectively, with $\beta = 4/5$. The resource levels are $r_l = 1$ and $r_h = 3/2$. Then $t^* = 1$ and $\bar{t} = 3/2$. Consider a group i with $q_i = 1/5$ and suppose that ability follows a symmetric beta distribution $a \sim Beta(x, x)$ with four possible specifications for the common shape parameter: $x \in \{1, 2, 5, 10\}$. Then $\hat{t}_i > t^*$ if $x \in \{1, 2\}$ and $\hat{t}_i = t^*$ otherwise.*

The example considers four possible specifications for the ability distribution, all of which belong to the family of symmetric beta distributions. The resulting four expected performance functions are shown in Figure 1. The plot on the top left shows how expected performance varies with training when ability is uniformly distributed (the shape parameters of the beta distribution both equal 1). In this case, expected performance is piecewise linear with a discontinuity at t^*. The figure on the top right has the same mean ability but lower variance (the shape parameters both equal 2). Here expected performance varies continuously with training but has a peak in the range $[0, t^*)$ at which expected performance is higher than at some training levels above t^*. The bottom panel shows ability distributions even more tightly clustered around the mean. In both cases there is non-monotonicity in the training range $[0, t^*)$ but performance above t^* always exceeds that at any training level below t^*.

Example 1 shows that the condition $\hat{t}_i > t^*$ may or may not be satisfied, depending on properties of the ability distribution. More generally, the condition also depends on the training and performance functions, the two resource levels, and the proportion of people with high resources. Since groups differ with respect to this last feature, it is also possible that this condition is satisfied for just one of the groups. Before examining the implications of this condition for non-monotonic selection, we establish the following result, which states that at any training level in the range $(0, t^*)$, the disadvantaged group must have higher expected performance.

Proposition 1. *$E(p_1|t) \geq E(p_2|t)$ at all training levels t, with strict inequality for $t \in (0, t^*)$.*

At training levels below t^* the disadvantaged group has higher expected performance because of a pool composition effect. As we show in the proof of the result, we must have $\gamma_1(t) < \gamma_2(t)$ as long as the density f is strictly positive at both $\alpha_h(t)$ and $\alpha_l(t)$.[9] Training levels above t^* can only be achieved with high resources, so candidates from both groups have the same performance in this range, and this performance is strictly increasing in training.

Proposition 1 immediately implies that if $\hat{t}_1 > t^*$, then $\hat{t}_1 > \hat{t}_2$ and $\hat{p}_1 > \hat{p}_2$. This is true regardless of whether we also have $\hat{t}_2 > t^*$. The following example illustrates, with p^* denoting the performance level of high resource individuals with training t^*.

Example 2. *Suppose that the two groups are of equal size and ability is distributed according to a symmetric beta distribution $a \sim Beta(2, 2)$. The training and performance functions are as in Example 1, and the shares with high resources in the two groups are $q_1 = 1/5$ and $q_2 = 2/3$. Then $(\hat{p}_1, \hat{p}_2, p^*) = (0.84, 0.76, 0.73)$ and $(\hat{t}_1, \hat{t}_2, t^*) = (1.14, 1.03, 1.00)$.*

Figure 2 shows how expected performance varies with training for the parameters in this example. At training levels below t^*, the disadvantaged group has higher expected performance, as claimed in Proposition 1.

We now define threshold levels of elite capacity that play a role in our analysis. Let $\hat{k}_i$ denote the proportion of the applicant population (aggregated across both groups) that has training at least $\hat{t}_i$:

$$\hat{k}_i = \sum_{i=1}^{2} s_i q_i (1 - F(\alpha_h(\hat{t}_i))). \tag{5}$$

Note that if $\hat{t}_1 > \hat{t}_2$, then $\hat{k}_1 < \hat{k}_2$. We then have:

[9] If ability is uniformly distributed, as in the first panel of Figure 1, then the disadvantaged group has higher performance also at t^*.

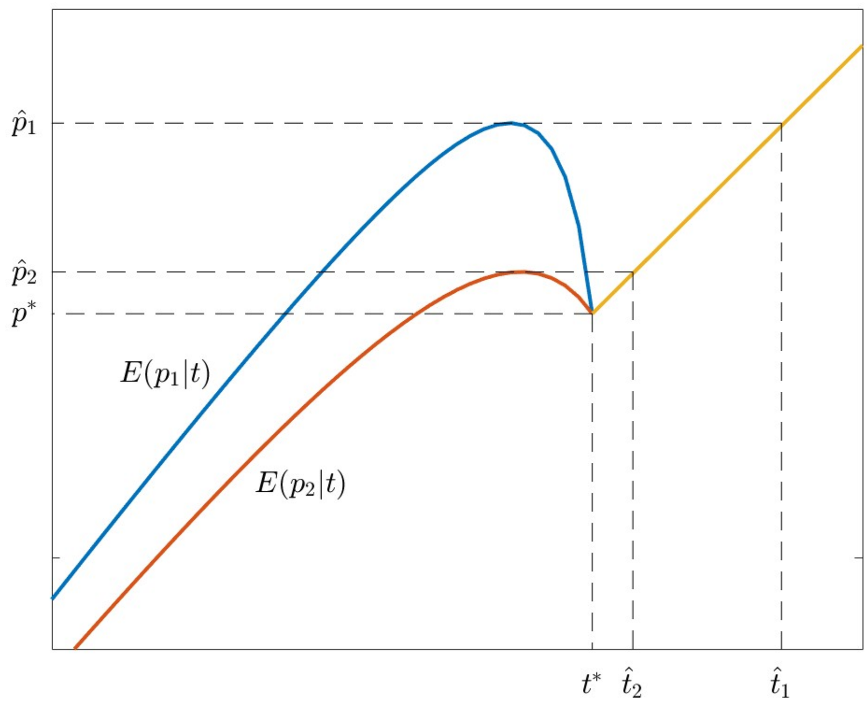

Figure 2: Expected performance conditional on training for two groups.

Proposition 2. *The performance-maximizing policy is monotonic and group-blind if* $k \leq \hat{k}_1$*, and is group-sighted and favors the disadvantaged group otherwise. Furthermore, if* $\hat{t}_1 > \hat{t}_2$ *and* $k \in (\hat{k}_1, \hat{k}_2)$*, then the disadvantaged group faces non-monotonic selection with some low resource individuals selected, while the advantaged group faces monotonic selection with only high resource individuals selected.*

Proposition 2 states that when elite capacity is at most $\hat{k}_1$, selection is monotonic and group blind. There is a common training threshold above which candidates from both groups are chosen, and only those with high resources are able to reach this threshold. For somewhat larger values of elite capacity in the range $(\hat{k}_1, \hat{k}_2)$ selection for the advantaged group is monotonic but the disadvantaged group faces non-monotonic selection. All those selected from the advantaged group have high resources, while some of those selected from the disadvantaged group have low resources.

To see why, note that when elite capacity k lies between $\hat{k}_1$ and $\hat{k}_2$, there is some performance level p between $\hat{p}_1$ and $\hat{p}_2$ such that the capacity k can be filled by accepting those with expected performance at least p and rejecting all others. When $\hat{p}_1$ is above p^* (as in Figure 2) then it must also be above $\hat{p}_2$, as a direct consequence of Proposition 1. In the advantaged group, those with expected performance at least p all have high resource access, and thus this group faces monotonic selection under the optimal policy. But in the disadvantaged group there are training levels below t^* at which expected performance exceeds p, and there are training levels above t^* at which expected performance lies below p. Hence the disadvantaged group faces non-monotonic selection under the optimal policy. This non-monotonicity creates incentives for some applicants to underinvest in training, and we consider the implications of this in Section 5.

It is clear that the disadvantaged group will be underrepresented under any monotonic and group-blind policy, and hence underrepresented when elite capacity is smaller than $\hat{k}_1$. However, further expansions in capacity could favor this group via a selection effect which mitigates their resource disadvantage. The size of the selection effect depends on the importance of ability relative to training in determining performance. We explore this next.

4.2 Representation

Define the *relative representation* ρ_i of group i as the fraction of individuals from this group who are selected, divided by the elite capacity k. Then, by definition,

$$s_1\rho_1 + s_2\rho_2 = 1.$$

If $\rho_1 = \rho_2 = 1$, we have proportional representation or statistical mirroring, and those in elite positions reflect the demographic composition of the population at large. The disadvantaged group is underrepresented if $\rho_1 < 1 < \rho_2$. More generally, the value of ρ_i ranges from zero (if no members are selected) to $1/s_i$ (if all members of group i and no members of the other group are selected). This can be very large if a small group occupies all elite positions.

Recall that when elite capacity $k < \hat{k}_1$, selection is monotonic and group-blind. There is a training threshold $t > \hat{t}_1$ above which all are selected and below which all are rejected, and only those with high resources can reach this threshold. In this case,

$$\rho_i = \frac{q_i(1 - F(\alpha_h(t))}{k} = \frac{q_i(1 - F(\alpha_h(t))}{(s_1q_1 + s_2q_2)(1 - F(\alpha_h(t))} = \frac{q_i}{s_1q_1 + s_2q_2}. \tag{6}$$

Note that this is independent of t (and hence of k). That is, when elite capacity is below $\hat{k}_1$, the relative representation of each group is constant. When the two groups are of equal size, this simplifies to $\rho_i = 2q_i/(q_1 + q_2)$, or twice the share of group i among high resource individuals in the population at large. For the specification in Example 2, the share of the disadvantaged group among high resource individuals is 3/13, and hence the relative representation of this group when k is sufficiently small is $6/13 \approx 0.46$. At the other extreme, as elite capacity becomes very large, the relative representation of both groups approaches 1, since almost all applicants are selected.

Between these two extremes, relative representation varies with elite capacity in interesting and complex ways. We explore this using the following parametric specification for the performance function:

$$p = \phi(a, t) = \beta a + (1 - \beta)t. \tag{7}$$

Here the parameter β captures the importance of ability relative to training in determining performance.

The training and capacity thresholds defined earlier will vary with β, and we write $\hat{t}_1(\beta)$ and $\hat{k}_1(\beta)$ to make this dependence explicit. Here $\hat{k}_1(\beta)$ is the level of elite capacity below which the performance-maximizing allocation is monotonic and group-blind, given a particular value of β. The following result states that higher values of β lower this threshold, and thus restrict the range of capacity levels over which allocations are monotonic and group blind:

Proposition 3. *Suppose that the performance function is given by $\phi(a, t) = \beta a + (1 - \beta)t$, and that $\hat{t}_1(\beta) > t^*$ for some $\beta \in (0, 1)$. Then $\beta' > \beta$ implies $\hat{k}_1(\beta') < \hat{k}_1(\beta)$.*

This result states that when the relative importance of ability rises in the performance function, the range of elite capacities for which the performance-maximizing allocation is monotonic and group-blind shrinks. This happens because the highest level of expected performance in the disadvantaged group rises with β. Since ability is exogenous and all individuals are achieving the highest feasible level of training, this is entirely a consequence of the fact that ability has to be inferred from training, and this inference favors the group with fewer high resource individuals. When ability matters more for performance, the effect of this informational feature on selection is amplified.

Proposition 3 may be illustrated using the following example.

Example 3. *Suppose that the two groups are of equal size and ability is uniformly distributed. The training and performance functions are as in Example 1, with $\beta \in \{0, 1/2, 5/6\}$. The shares of the population with high resources in the two groups are $q_1 = 1/5$ and $q_2 = 1/2$ respectively. Then higher values of β correspond to lower values of $\hat{k}_1$, and greater representation of the disadvantaged group when $k > \hat{k}_1$ (see Figure 3).*

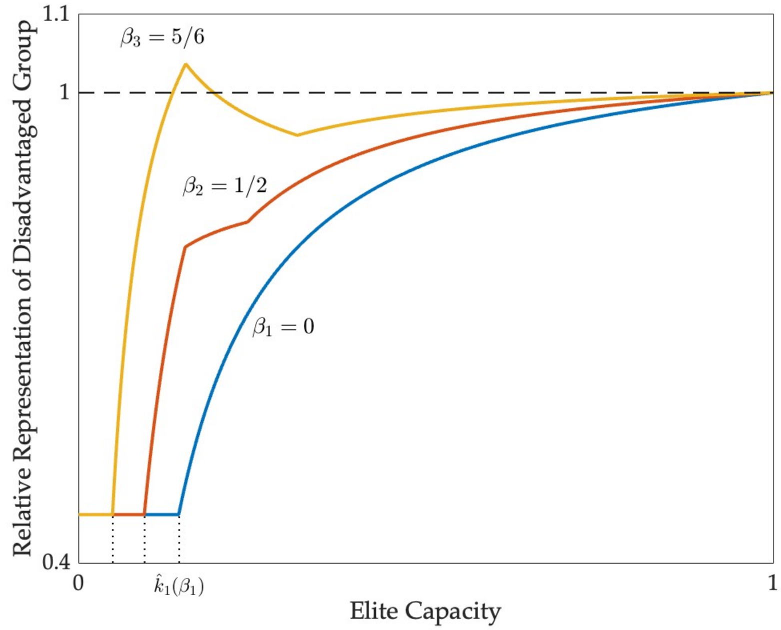

Figure 3: Relative representation of the disadvantaged group for different values of β.

Figure 3 traces the relationship between relative representation of the resource disadvantaged group and elite capacity for three alternative values of β, based on the specifications in this example. The horizontal dashed line in the figure corresponds to statistical mirroring, where the same proportion k from each group is selected. The lowest curve corresponds to $\beta = 0$, so only training matters for productivity. This results in performance maximizing policies being both monotonic and group-blind, with the disadvantaged group always underrepresented. When ability also has an independent effect on productivity the disadvantaged group benefits from the screening effect, to a degree that is increasing in β. For high enough values of β, there is a range of capacity levels for which the meritocratic allocation results in *overrepresentation* of the disadvantaged group because the screening effect is strong enough to overcome the resource disadvantage for the selected applicants.

5 Strategic Investment

To this point we have assumed that all applicants attain the highest training level that is feasible for them given their ability and resources. However, when selection policies are non-monotonic, some applicants face incentives to underinvest in training in order to raise their prospects of selection. Since performance is increasing in training (conditional on ability), underinvestment degrades the quality of the applicant pool and is therefore costly for the organization.

The organization can avoid underinvestment by committing to a selection policy that is weakly monotonic in training, i.e. $\pi_i(t)$ is non-decreasing in t for each group i.[10] Such policies will not be deterministic in general—there may be training levels t at which $\pi_i(t)$ lies strictly between zero and one for one or both groups. However, as long as $\pi_i(t)$ is weakly monotonic for each group, there will be no incentives to underinvest in training, and we assume that all applicants attain the highest level of training consistent with their ability and resources. We now explore the effects of such policies on optimal selection and representation.

If elite capacity is below $\hat{k}_1$, we have already shown that performance is maximized with a monotonic and group-blind policy in the absence of strategic training choices. This policy therefore remains optimal when allowing for strategic behavior. Our focus in this section is on characterizing optimal policies for capacity levels above $\hat{k}_1$. We now show that these could result in either higher or lower representation of the disadvantaged group relative to our benchmark case of deterministic training.

Before turning to details, note that the organization is choosing between two broad types of policies. It can either choose to forego hiring the applicants with the highest expected perfromance but lower training because it does not want to be forced to hire those with low expected performance at intermediate training levels. This would lead to monotonic and group-blind policies that would favor the advantaged group. Alternatively, it could accept the highest performance applicants with positive probability but be forced to also accept some with higher training but lower performance with the same probability. For reasons discussed below, this could favor the disadvantaged group to an even greater degree than would be the case under unconstrained choices. Which of these policy regimes is optimal depends on the level of elite capacity and other parameters of the model, as we now show.

First consider the case of elite capacity slightly above $\hat{k}_1$. Recall that $\hat{p}_1$ and $\hat{p}_2$ are

[10] Without commitment the equilibrium policy may fail even weak monotonicity. An example of this for a discrete version of the model can be found in Sethi and Somanathan (2023). Policy induced declines in human capital investment are also a feature of the Coate and Loury (1993) model, but for different reasons—applicants reduce costly investment in skills because they face a lower selection threshold under a monotonic rule.

the highest levels of expected training, for the disadvantaged and advantaged group respectively, at training levels below t^* and p^* is expected performance at t^*. Suppose that $\hat{p}_1 > p^*$ (as in Figure 2). In this case, the unconstrained decision-maker would select some applicants with training below t^* and performance close to $\hat{p}_1$, resulting in non-monotonic selection for applicants from the disadvantaged group. Such a policy cannot be part of an equilibrium with strategic underinvestment, since all those who are passed over would choose lower training levels, bringing down expected performance and causing capacity to be exceeded. As a result, when elite capacity is slightly above $\hat{k}_1$, the optimal policy will be strictly monotonic and group blind, with a common training threshold slightly below $\hat{t}_1$ for all applicants. At these levels of elite capacity, the possibility of strategic underinvestment lowers the representation of the disadvantaged group.

At some higher levels of elite capacity, we show that allowing for strategic training choices can *increase* the representation of the disadvantaged group. Suppose that the inequalities $p^* \leq \hat{p}_2 \leq \hat{p}_1$ both hold strictly, as in Figure 2. Then the set $P = [p^*, \hat{p}_1]$ has positive measure. For any $p \in P$, define the training levels $t_l(p)$ and $t_h(p)$ as the lowest and highest training levels at which the expected performance of disadvantaged group members is p. That is:

$$\begin{aligned} t_l(p) &= \min\{t | E(p_1|t) = p\} \\ t_h(p) &= \max\{t | E(p_1|t) = p\} \end{aligned}$$

Since $t_h(p) > t^*$, only high resource individuals can reach this training level, so it is given by the unique solution to

$$\phi(\alpha_h(t), t) = p.$$

A particular training level in P plays a key role in our analysis and is identified in the following lemma:

Lemma 1. *There exists a training level $p^e \in P$ such that*

$$E(p_1 | t_l(p^e) \leq t \leq t_h(p^e)) = p^e,$$

and no higher training level satisfies this condition.

That is, the expected performance of all those in the disadvantaged group with training in the interval $[t_l(p^e), t_h(p^e)]$ is precisely equal to p^e. If there are multiple such training levels, p^e simply corresponds to the largest of them. We then have the following result on representation of the disadvantaged group when strategic investments in training are possible:

Proposition 4. *Suppose that $p^e > \hat{p}_2$. Then there exists an interval K of elite capacities such that if $k \in K$, the representation of the disadvantaged group is greater when strategic underinvestment is possible than when it is not.*

To get some intuition for this result, let k_l denote the share of applicants, aggregating across both groups, with training above $t_h(p^e)$, and let k_h be the share obtained by adding to k_l the applicants from the disadvantaged group with training between $t_l(p^e)$ and $t_h(p^e)$. When elite capacity lies between these two thresholds, the optimal selection policy has the following structure: all those with training at least $t_h(p^e)$ (regardless of group) are selected with certainty, and all those in the disadvantaged group with training between $t_l(p^e)$ and $t_h(p^e)$ are selected with positive probability. This probability rises from zero to one as elite capacity rises from k_l to k_h. When it is equal to k_l the advantaged group is favored (relative to the unconstrained case) and when it is equal to k_h the disadvantaged group is favored. The interval K in the result is accordingly a subset of $(k_l, k_h]$. This subset is characterized in the appendix.

Thus, at some levels of elite capacity, the underrepresentation of the disadvantaged group is more severe when applicant choices are strategic than when they are not. Meanwhile, provided that $p^e > \hat{p}_2$, the opposite is true at other levels of elite capacity. The possibility of strategic underinvestment forces admissions policies to be weakly monotonic, which makes it difficult to exclude applicants at training levels that would otherwise indicate low expected performance. The decision-maker can respond to this by excluding groups that have lower expected performance at those training levels. That is, the possibility of strategic applicant choices does not overturn the results obtained in Section 4. In fact, the effects identified there can be amplified.

6 Discussion

Any practical implementation of a meritocratic ideal must be based on observable credentials. In popular discourse, merit-based allocations are inconsistent with selection based on markers of group identity, such as ethnicity, gender, or religion, since these are not intrinsically related to performance. It is also believed that selection based on merit ought to satisfy a monotonicity property; those with higher values of performance-related attributes, such as test scores or course grades, should have precedence over those with lower values. In this paper we have challenged the general nature of both of these claims.

Our arguments reveal that the usual framing of the problem of affirmative action—as a trade-off between performance and representation—is misleading.

We show that selection rules that maximize organizational performance can be both group-conscious and non-monotonic in observable credentials. Whether or not they are in fact so, depends on the joint distribution of observable and unobservable attributes (which we term training and ability respectively), their relative importance in determining performance, and the fraction of the eligible pool of candidates that is selected. Without these pieces of information, we cannot reasonably infer the weights (if any) that organizations attach to diversity based on the degree of representation among their students or workers.

We also show that when observable signals of performance, such as training in our model, can be strategically manipulated, non-monotonic selection will not be observed in equilibrium, since applicants that are rejected could pool with those with lower levels of training. We show, surprisingly, that organizational responses to such strategic behavior can, for some levels of capacity, increase the representation of disadvantaged groups.

An implication of our results is that the degree of representation in market allocations is likely to vary across sectors, and across institutions within each sector. For example, within a university system, the elite colleges may admit those with both high resources and high ability, while second tier institutions select those with high ability and low resources. The latter set of institutions would have greater group representation even though both institutions are pure performance maximizers. It is not therefore possible to infer the extent to which institutions have specifically pursued representation goals simply by looking at the composition of students they have admitted. To assess the extent to which representation goals are explicitly pursued, one would have to solve for the market equilibrium in the case where all maximize expected performance and then consider deviations in representation from this allocation.

Our results can also help evaluate quota policies that are used in many countries. Many of these use statistical mirroring, whereby the fraction of seats allocated to disadvantaged minorities is equal to their share in the population. We show while in some cases these may be more efficient than a single standard for all groups, there are gains in efficiency from allowing for different degrees of representation across institutions while maintaining overall representation targets. We hope our model provides a framework for future research in this direction.

Our model involves selection into positions that are considered desirable by applicants, who are chosen based on their expected performance. An interesting extension would involve positions that people seek to avoid, such as arrest or incarceration. In this case those with greater resources may be able to exhibit a broader range of observable characteristics than those with more limited resources. Resource inequality across groups would then imply overrepresentation of the disadvantaged group in undesirable positions under monotonic and group-blind selection. But selection will not be monotonic and group-blind in general, and questions related to the manner in which representation varies with capacity may be worth exploring.

A Appendix

A.1 Proofs

Proof of Proposition 1. Note that for any $t' \leq t \leq t^*$,

$$\begin{aligned}\Pr[t' \leq t | r = r_l] &= F(\alpha_l(t)) \\ \Pr[t' \leq t | r = r_h] &= F(\alpha_h(t))\end{aligned}$$

These are the distribution functions of training for low and high resource individuals respectively. The corresponding density functions are $f(\alpha_l(t))\alpha_l'(t)$ and $f(\alpha_h(t))\alpha_h'(t)$. Using Bayes' rule, when $t \leq t^*$, the probability that an individual in group i has high resources conditional on training t is given by

$$\gamma_i(t) = \frac{q_i f(\alpha_h(t))\alpha_h'(t)}{q_i f(\alpha_h(t))\alpha_h'(t) + (1-q_i) f(\alpha_l(t))\alpha_l'(t)}. \tag{8}$$

It is easily verified that $q_1 < q_2$ implies $\gamma_1(t) < \gamma_2(t)$ for any $t \in (0, t^*)$. Recall from (3) that expected performance is defined as

$$E(p_i|t) = \gamma_i(t)\phi(\alpha_h(t), t) + (1 - \gamma_i(t))\phi(\alpha_l(t), t).$$

Since $\phi(\alpha_h(t), t) < \phi(\alpha_l(t), t)$, we therefore obtain

$$E(p_1|t) > E(p_2|t)$$

as claimed. It is easily verified that since only high resource individuals reach training levels $t > t^*$, we have $E(p_1|t) = E(p_2|t)$ for training in this range. □

Proof of Proposition 2. Define $t_m(k)$ as the training level such that a proportion k of applicants exceed it. That is, if elite capacity is k then a monotonic and group-blind policy with threshold $t_m(k)$ will exactly meet it. Note that $t_m(k)$ is strictly decreasing and $t_m(\hat{k}_i) = t_i$ for each group i. If $k \leq k_1$ then $t_m(k) \geq t_1$. Suppose this is the case. Then for any t_h, t_l satisfying $t_l < t_m(k) \leq t_h$, and any $i, j \in \{1, 2\}$, we have

$$E(p_i|t_h) > E(p_j|t_l).$$

That is, all those selected under a monotonic and group-blind policy have higher expected performance than any of those rejected. Hence the policy is optimal.

Now suppose that $k > \hat{k}_1$, so $t_m(k) < \hat{t}_1$. If $t_m(k) < t^*$ then a monotonic and group blind-policy must select some applicants with training below t^*. By Proposition 1, any optimal policy under which candidates with training below t^* are selected cannot be group-blind. Hence the optimal policy cannot be group-blind if $t_m(k) < t^*$. If, instead, $t_m(k) \in [t^*, \hat{t}_1)$, then by the definition of $\hat{t}_1$, there exists a training level $t < t^*$ such that

$$E(p_1|t) > E(p_1|t_m(k)).$$

This again implies from Proposition 1 that the optimal policy cannot be group blind. Proposition 1 also implies that any optimal group-sighted policy must favor the disadvantaged group, which proves the first claim.

To prove the second claim, suppose that if $\hat{t}_1 > \hat{t}_2$ and $k \in (\hat{k}_1, \hat{k}_2)$. In this case $t_m(k) \in (\hat{t}_2, \hat{t}_1) \subset [t^*, \hat{t}_1)$. As shown above, the optimal policy in this case must be group-sighted and favor the disadvantaged group. Furthermore, since the monotonic and group-blind policy at this level of elite capacity ensures that all selected applicants have expected performance exceeding $\hat{p}_2$, this must also be true of the optimal policy. There are no members of the advantaged group with training below $\hat{t}^2$ that can reach this level of expected performance. Hence all those selected from the advantaged group have training at least $\hat{t}_2$, and thus have high resource access. □

Proof of Proposition 3. Consider any $\beta \in (0, 1)$ such that $\hat{t}_1(\beta) > t^*$. Let $\tilde{t}_1(\beta)$ denote the highest training level in the range $[0, t^*]$ at which the expected performance in group 1 is maximized, and hence equal to the performance at $\hat{t}_1$. Hence

$$E(p_1|\tilde{t}_1) = \gamma_1(\tilde{t}_1)\phi(\alpha_h(\tilde{t}_1), \tilde{t}_1) + (1 - \gamma_1(\tilde{t}_1))\phi(\alpha_l(\tilde{t}_1), \tilde{t}_1) = \phi(\alpha_h(\hat{t}_1), \hat{t}_1),$$

where the dependence of $\tilde{t}_1$ and $\hat{t}_1$ on β has been supressed for expositional clarity. Using the specification in (7) for the performance function, we therefore have

$$\gamma_1(\tilde{t}_1)(\beta\alpha_h(\tilde{t}_1) + (1 - \beta)\tilde{t}_1) + (1 - \gamma_1(\tilde{t}_1))(\beta\alpha_l(\tilde{t}_1) + (1 - \beta)\tilde{t}_1)) = \beta\alpha_h(\hat{t}_1) + (1 - \beta)\hat{t}_1.$$

Equivalently,

$$\beta(\gamma_1(\tilde{t}_1)\alpha_h(\tilde{t}_1) + (1 - \gamma_1(\tilde{t}_1))\alpha_l(\tilde{t}_1)) + (1 - \beta)\tilde{t}_1 = \beta\alpha_h(\hat{t}_1) + (1 - \beta)\hat{t}_1.$$

Since $\tilde{t}_1 < \hat{t}_1$, this implies

$$\gamma_1(\tilde{t}_1)\alpha_h(\tilde{t}_1) + (1 - \gamma_1(\tilde{t}_1))\alpha_l(\tilde{t}_1) > \alpha_h(\hat{t}_1).$$

Hence, for any $\beta' > \beta$, we have

$$\beta'(\gamma_1(\tilde{t}_1)\alpha_h(\tilde{t}_1) + (1-\gamma_1(\tilde{t}_1))\alpha_l(\tilde{t}_1)) + (1-\beta')\tilde{t}_1 > \beta'\alpha_h(\hat{t}_1) + (1-\beta')\hat{t}_1.$$

Note that the thresholds $\tilde{t}_1$ and $\hat{t}_1$ correspond to β, not β'. Hence we have

$$E(p_1|\tilde{t}_1(\beta)) > E(p_1|\hat{t}_1(\beta))$$

when the weight on ability in the performance function is β'. This implies $\hat{t}_1(\beta') > \hat{t}_1(\beta)$, and hence $\hat{k}_1(\beta') < \hat{k}_1(\beta)$ as claimed. □

Proof of Lemma 1. Define the continuous function $\lambda : P \to P$ as follows:

$$\lambda(p) = E(p_1|t_l(p) \leq t \leq t_h(p)).$$

This is the expected performance among those in the disadvantaged group with training levels between $t_l(p)$ and $t_h(p)$. Note that $\lambda(\hat{p}_1) < \hat{p}_1$ and $\lambda(p^*) > p^*$. Hence there exists at least one solution in P to $\lambda(p) = p$. Let p^e denote the largest of these solutions. □

Proof of Proposition 4. Define two threshold elite capacity levels as follows:

$$\begin{aligned} k_l &= s_2q_2(1-F(\alpha_h(t_h(p^e)))) + s_1q_1(1-F(\alpha_h(t_h(p^e)))) \\ k_h &= s_2q_2(1-F(\alpha_h(t_h(p^e)))) + s_1q_1(1-F(\alpha_h(t_l(p^e)))) + s_1(1-q_1)(1-F(\alpha_l(t_l(p^e)))), \end{aligned}$$

where p^e is as defined in Lemma 1. Here k_l is the population (aggregating across both groups) with training above $t_h(p^e)$ and k_h adds to this the population of the disadvantaged group with training between $t_l(p^e)$ and $t_h(p^e)$.

Now suppose that $p^e > \hat{p}_2$ and $k \in (k_l, k_h]$. Then there exists $\eta \in (0, 1]$ such that the following selection policy is weakly monotonic and feasible:

$$\pi_1(t) = \begin{cases} 1 & \text{if } t \geq t_h(p^e) \\ \eta & \text{if } t_l(p^e) \leq t < t_h(p^e) \\ 0 & \text{otherwise} \end{cases}$$

and

$$\pi_2(t) = \begin{cases} 1 & \text{if } t \geq t_h(p^e) \\ 0 & \text{otherwise} \end{cases}$$

Call this policy $\mathcal{P}$. We first show that given the weak monotonicity constraint on policies, $\mathcal{P}$ results in higher expected performance than any alternative under which applicants rejected under $\mathcal{P}$ are accepted. To see this, note that all applicants rejected under $\mathcal{P}$ have performance at most p^e. If some of these are to be accepted, then they must replace applicants from the disadvantaged group with performance below p^e who are accepted under $\mathcal{P}$, for example those with training between t^* and $t_h(p^e)$. But any decline in the probability of acceptance η for these candidates must also result in an equal or greater decline in the probability of acceptance for all candidates under $\mathcal{P}$ to maintain weak monotonicity. This pool of candidates, taken together, has expected performance at least p^e, and replacing them would lower expected performance.

Thus $\mathcal{P}$ is superior to any policy under which candidates rejected under $\mathcal{P}$ are accepted. To complete the proof, we show that $\mathcal{P}$ has greater representation of the disadvantaged group than the unconstrained policy when k is sufficiently close to k_h. This is clear for $k = k_h$ since the unconstrained policy would involve some performance threshold p between p^* and p^e above which all applicants are accepted, regardless of group. This would replace some disadvantaged group members accepted under the constrained policy with an equal measure of advantaged group members who are rejected. It is easily verified that this would also hold for $k = k_h - \varepsilon$ for ε sufficiently small. Thus there exists $K \subset (k_l, k_h]$ such that the disadvantaged group has higher representation when applicants are strategic than when they are not. □

A.2 Continuous Resource Distributions

While the case of two resource levels makes our reasoning transparent, this is not required to generate either the non-monotonicity of expected performance in training, or selection policies that favor the disadvantaged group. This can be seen by considering a case with continuous resource distributions.

Suppose that ability and resources are independently distributed on $[0, 1]$, with the ability distribution being uniform. The two resource distributions are both beta distributions with shape parameters (2,10) in the disadvantaged group and (10,2) in the advantaged group. The training function is $\tau(a, r) = ar$ and there is no strategic underinvestment. The two groups are of equal size.

Conditional on training t, the expected ability in each group, as well as that in the population as a whole, is shown in Figure 4. The disadvantaged group has higher expected ability (conditional on training) at all training levels, and hence also has higher expected performance conditional on training, regardless of the performance function (as long as performance depends on both ability and training). So the disadvantaged group will be

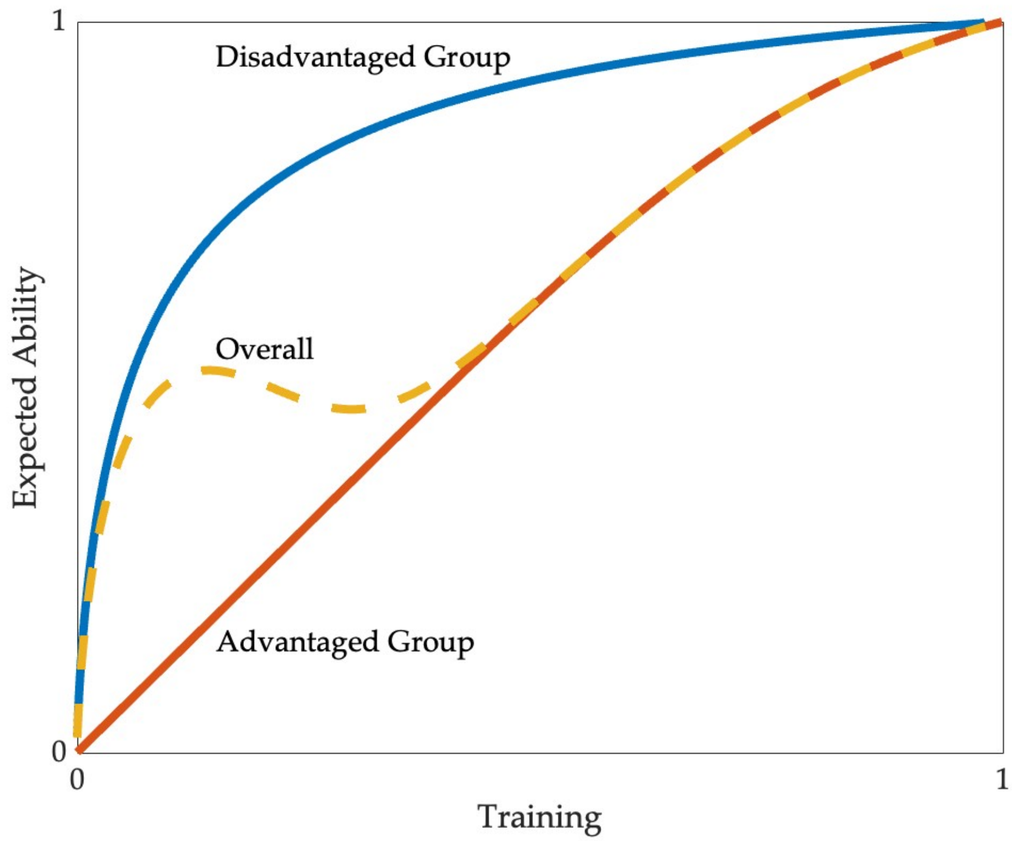

Figure 4: Expected ability by group and overall.

favored under a performance-maximizing selection policy (regardless of elite capacity) and the imposition of group-blind selection will lower performance.

Although expected ability is increasing in training within any group, this is not the case for the population as a whole. Whether or not expected performance in the population as a whole is likewise non-monotonic will depend on the degree to which ability rather than training matters for performance. If ability has sufficient weight, the overall expected performance will also inherit the non-monotonicity of expected ability. In this case the imposition of a group-blind policy will result in non-monotonic selection to maximize performance.

This example also shows that expected performance can be non-monotonic *within* groups. For instance, the overall population here can itself be considered as a single group, with a bimodal resource distribution obtained by simply aggregating the two component distributions.

References

Dennis J. Aigner and Glen G. Cain. Statistical theories of discrimination in labor markets. Industrial and Labor Relations Review, pages 175–187, 1977.

Richard Arneson. Equality of Opportunity. In Edward N. Zalta, editor, The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Summer 2015 edition, 2015.

Joshua Aronson and Claude M. Steele. Stereotypes and the fragility of academic competence, motivation, and self-concept. Handbook of competence and motivation, pages 436–456, 2005.

Kenneth J. Arrow. The theory of discrimination. In Orley Ashenfelter and Albert Rees, editors, Discrimination in Labor Markets. Princeton, NJ: Princeton University Press, 1973.

Ian Ball. Scoring strategic agents. American Economic Journal: Microeconomics, forthcoming, 2024.

Matt Cavanagh. Against Equality of Opportunity. Clarendon Press, 2002.

Dario Cestau, Dennis Epple, and Holger Sieg. Admitting students to selective education programs: Merit, profiling, and affirmative action. Journal of Political Economy, 125 (3):761–797, 2017.

Jimmy Chan and Erik Eyster. Does banning affirmative action lower college student quality? American Economic Review, 93(3):858–872, 2003.

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

Bradford Cornell and Ivo Welch. Culture, information, and screening discrimination. Journal of Political Economy, pages 542–571, 1996.

Wouter Dessein, Alex Frankel, and Navin Kartik. Test-optional admissions. arXiv preprint arXiv:2304.07551, 2023.

Steven N. Durlauf. Affirmative action, meritocracy, and efficiency. Politics, Philosophy & Economics, 7(2):131–158, 2008.

Alex Frankel and Navin Kartik. Muddled information. Journal of Political Economy, 127 (4):1739–1776, 2019.

Alex Frankel and Navin Kartik. Improving information from manipulable data. Journal of the European Economic Association, 20(1):79–115, 2022.

Roland G. Fryer and Glenn C. Loury. Valuing diversity. Journal of Political Economy, 121 (4):747–774, 2013.

Roland G. Fryer, Glenn C. Loury, and Tolga Yuret. An economic analysis of color-blind affirmative action. Journal of Law, Economics, and Organization, 24(2):319–355, 2008.

Carnes Lord. Aristotle's Politics, Second Edition. University of Chicago Press, 2013.

Daniel Markovits. The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite. Penguin Press, 2019.

Edmund S. Phelps. The statistical theory of racism and sexism. American Economic Review, 62(4):659–661, 1972.

Debraj Ray and Rajiv Sethi. A remark on color-blind affirmative action. Journal of Public Economic Theory, 12(3):399–406, 2010.

John E Roemer. Equality of Opportunity. Harvard University Press, 1998.

Michael J Sandel. The Tyranny of Merit: What's Become of the Common Good? Farrar, Straus and Giroux, 2020.

TM Scanlon. Why Does Inequality Matter? Oxford University Press, 2018.

Amartya Sen. Merit and justice. In Kenneth Arrow, Samuel Bowles, and Steven N Durlauf, editors, Meritocracy and Economic Inequality, pages 5–16. Princeton University Press, 2000.

Rajiv Sethi and Rohini Somanathan. Meritocracy and representation. Journal of Economic Literature, 61(3):941–957, 2023.

Michael Spence. Job market signaling. The Quarterly Journal of Economics, 87(3):355–374, 1973.

Gregory M Walton, Steven J Spencer, and Sam Erman. Affirmative meritocracy. Social Issues and Policy Review, 7(1):1–35, 2013.