AFFIRMATIVE ACTION AND STEREOTYPES IN HIGHER EDUCATION ADMISSIONS*

Prasad Krishnamurthy
U.C. Berkeley Law School
prasad@law.berkeley.edu

Aaron Edlin
U.C. Berkeley Dept. of Economics
U.C. Berkeley Law School
National Bureau of Economic Research
edlin@berkeley.edu

This version: October 13, 2014

Abstract

We analyze how admission policies affect stereotypes against students from disadvantaged groups. Many critics of affirmative action argue that lower admission standards cause such stereotypes and suggest group-blind admissions as a remedy. We show that when stereotypes result from social inequality, they typically persist under group-blind admissions. Perversely, eliminating stereotypes requires a higher admission standard for disadvantaged students. If a school seeks both to treat students equally and limit stereotypes, the optimal admission policy would still impose a higher standard on disadvantaged students. A third goal, such as equal representation, is required to justify group-blind admissions. Even in this case, group-blind admissions are optimal only when the conflicting goals of equal representation and limiting stereotypes exactly balance. This is an implausible justification for group-blind admission because it implies that some schools desire higher standards for disadvantaged students. Therefore, if a school values all three of these goals, some amount of affirmative action will be optimal.

*Berkeley Law and Economics Working Paper. Please do not cite without the authors' permission. We thank Bobby Bartlett, Omri Ben-Shahar, Richard Brooks, Bob Cooter, Dhammika Dharmapala, Lee Fennell, David Gamage, Jonah Gelbach, Mark Gergen, Bert Huang, Will Hubbard, Saul Levmore, Yair Listokin, Bentley MacLeod, Richard McAdams, Justin McCrary, Martha Nussbaum, David Oppenheimer, Vicki Plaut, Ariel Porat, Eric Posner, Kevin Quinn, Russell Robinson, Arden Rowell, Alex Stremitzer, Talha Syed, Eric Talley, David Weisbach, Glenn Weyl, and participants at the NBER Law and Economics Summer Institute, University of Chicago Law and Economics Workshop, Berkeley Law Faculty Workshop, and the University of Illinois Law and Social Science Workshop for helpful comments.

"When blacks take positions in the highest places of government, industry, or academia, it is an open question today whether their skin color played a part in their advancement. The question itself is the stigma–because either racial discrimination did play a role, in which case the person may be deemed 'otherwise unqualified,' or it did not, in which case asking the question itself unfairly marks those who would succeed without discrimination."

Justice Clarence Thomas

"I hear the stigma argument all the time. 'But affirmative action causes stigma.' Well, yes, affirmative action causes stigma. That's one of the costs of affirmative action, I acknowledge it. It's a cost worth paying."

Christopher Edley

Most critics of affirmative action argue that it is discriminatory. Many also argue that affirmative action harms its intended beneficiaries. As Justice Clarence Thomas states in his concurring opinion in *Fisher v. Texas*, "There can be no doubt that the University's discrimination injures white and Asian applicants who are denied admission because of their race. But I believe the injury to those admitted under the University's discriminatory admissions program is even more harmful." For Justice Thomas and other critics, one reason that affirmative action is harmful is because it results in the stereotype that its beneficiaries are less qualified or able than others.[1]

In this paper, we analyze how admission policies affect stereotypes through statistical discrimination against students from disadvantaged groups. Contrary to what some critics of affirmative action suggest, a concern for the effects of stereotypes does not imply that a school should adopt group-blind admissions. We show that when stereotypes are a result of social disadvantage, they typically persist even if schools adopt group-blind admissions. Eliminating stereotypes requires a school to adopt *higher* admissions standards for students from disadvantaged groups. Such a perverse double standard is clearly unacceptable. If stereotypes are a concern, the appropriate question to ask is not whether they can be eliminated, but how much admission policy should target stereotypes relative to pursuing other goals.

We present a simple model to illustrate the relationship between admission policies and stereotypes. A school chooses to admit students from an advantaged and a disadvantaged group based on an academic score. The disadvantaged group has a lower average score

[1] In response to this view, a large literature in social science attempts to measure whether affirmative action programs have an effect on the perceived competence of beneficiaries (Heilman et al (1992 and 1997), Nye (1998), Evans (2003)). A related literature examines whether stereotypes affects student performance (Steele (1997), Steele and Aronson (1998), Fischer and Massey (2006)), which would compound the injury.

than the advantaged group.[2] We define a negative stereotype as statistical discrimination (Arrow (1972), Phelps (1972), Coate and Loury (1993)) against the disadvantaged group on the basis of this score. This definition is consistent with how test scores and other measures of academic performance are used in many discussions of affirmative action. For example, in *Fisher v. Texas* Justice Thomas criticizes the effects of affirmative action by pointing out that among students admitted outside of the University of Texas's Top Ten Percent Plan, "Blacks had a mean GPA of 2.57 and a mean SAT score of 1524; Hispanics had a mean GPA of 2.83 and a mean SAT score of 1794; whites had a mean GPA of 3.04 and a mean SAT score of 1914; and Asians had a mean GPA of 3.07 and a mean SAT score of 1991."[3]

We first demonstrate that negative stereotypes can persist even under group-blind admissions. For example, if all students above a given score are admitted, there may be few disadvantaged students with very high scores and many with scores just above the admissions cutoff. As a result, admitted students from the disadvantaged group may still have a lower average score than other admitted students. We show that a negative stereotype will exist under any group-blind admission policy if the distributions of scores for the groups satisfy the monotone likelihood ratio property. To illustrate the plausibility of this property, we show that it holds in a representative sample of ACT scores of high school students who self identify as "black" and "white."[4] Therefore, racial stereotypes would continue to exist under any race-blind admission policies based on the ACT or similar tests.

Perversely, eliminating a negative stereotype requires actively discriminating against disadvantaged students. We show that to eliminate stereotypes a school must make it more difficult for disadvantaged students to gain admission than for advantaged students. As a result, the percentage of admitted students who are disadvantaged will be even lower than the percentage under group-blind admissions. The greater the inequality between the two groups, the greater the admission penalty that must be imposed on disadvantaged students to eliminate any stereotype. Using ACT data, we show that eliminating the racial stereotype through admission policy would require a substantial penalty on black students and greatly reduce their representation.

An admission policy that would eliminate stereotypes is, admittedly, extreme. We therefore ask if other goals in admissions change this extreme result. We show that if a

[2] In particular, we assume that the distribution of scores from the advantaged group first-order stochastically dominates the distribution from the disadvantaged group. This definition reflects the view that deeper social forces give rise to the inequality between the two groups. For example, a student from a disadvantaged group will have fewer opportunities to acquire many of the characteristics valued by schools in admissions.

[3] These figures are the SAT scores and freshman-year GPAs for freshmen who entered the University of Texas at Austin in 2009.

[4] The ACT does not entirely determine any school's admission policy, but tests like the SAT and ACT do play a substantial role in admissions for many selective colleges (Krueger, Rothstein, and Turner (2006)).

school is willing to consider tradeoffs between equal treatment and combating sterotypes, it is still optimal to set a higher admission standard for disadvantaged students. We define equal treatment as the desire to treat all applicants the same, regardless of group status,[5] and consider the choice of admission policy by a school with preferences over both equal treatment and limiting stereotypes. When both these ends are valued with diminishing margins, it is never optimal to choose a group-blind admission policy.[6]

In fact it requires a third goal, such as equal representation, to justify group-blind admissions. We define equal representation as the desire for the demographics of admitted students to reflect those of the broader population and consider the choice of admission policy by a school with preferences over equal treatment, combating stereotypes, and equal representation. We show that group-blind admissions are optimal only when concerns over equal representation and reducing stereotypes exactly balance. Because this is a knife-edge case, it is an implausible justification for group-blind admissions. Given any heterogeneity among schools, if some schools are in perfect balance, this implies that other schools should actually desire higher standards for disadvantaged students. No school, however, would support such a discriminatory double standard. Therefore, some amount of affirmative action is typically optimal (except in boundary cases) if a school values all three of these goals with diminishing margins.

We conclude that harm from stereotypes contributes little to the case for group-blind admissions. Many critics of affirmative action advocate group-blind admissions in all circumstances. According to our analysis, this view is best described by a lexicographic preference for equal treatment. Such preferences are actually suggested implicitly by Justice Thomas in his concurring opinion in *Fisher v. Texas*, when he states that "for constitutional purposes" it does not matter whether the effects of affirmative action are "insidious" or "benign."

Our conclusions follow when stereotypes are based on differences in the distributions of characteristics in two admitted-student populations, and do not necessarily apply to all forms of stereotyping. In particular, our conclusions do not apply to "stereotypes" that only exist in the presence of affirmative action. For example, if a negative stereotype is defined as the probability that a student is admitted because of affirmative action, then ending affirmative action would eliminate the stereotype.

The effect of admission standards on group stereotypes is one of many concerns in formulating an admission policy. Affirmative action will not limit such stereotypes—in the way we define them—to the same extent as group-blind admissions. As Christopher

[5] This formal or procedural definition tracks the way equality is described by critics of affirmative action.

[6] Throughout, we make the distinction between (i) a preference for equal treatment in admissions that may exist alongside other preferences, and (ii) the choice of a group-blind admission policy. We try to be clear in the text when we are referring to a preference or a choice.

Edley suggests, stereotypes may be a necessary cost of affirmative action. However, when stereotypes result from disadvantage they are a likely feature of any admission policy, except perverse ones that discriminate against the disadvantaged. The issue for admission policy is therefore not whether it can eliminate stereotypes, but how much it should seek to affect stereotypes relative to pursuing other goals.

This paper contributes to the literature analyzing affirmative action and statistical discrimination (Fang and Moro (2011)). It is closely related to Coate and Loury (1993), who explain how racial stereotypes can be self-fulfilling as a result of the strategic interaction between workers and firms. Unlike Coate and Loury (1993), we focus on the decision problem of a school and ignore the effort choice of students so as to consider different values in admissions. It is also related to Chan and Eyster (2003), who analyze optimal admission policies when schools value racial diversity and student quality, but are constrained from explicitly using race. Other papers analyzing optimal admissions under similar constraints include Fryer, Loury, and Yuret (2008), Chan and Eyster (2009), and Ray and Sethi (2010). Section I describes the relationship between admission policy and stereotypes. Section II describes admission policy when schools also value equal treatment and equal representation. Section III discusses how our definition of stereotype is related to other definitions.

I Admission Policy and Stereotypes

A A Simple Model of Admissions

Consider a single school that accepts students under competitive admission. The admission pool consists of advantaged (A) and disadvantaged (D) groups of applicants, where θ_A and θ_D represent the fraction of each group in the pool ($\theta_A + \theta_D = 1$). The scores for an individual from each group are random variables s_A and s_D distributed on $[\underline{s}, \overline{s}]$ according to $F_A(s)$ and $F_D(s)$ with associated probability density functions $f_A(s)$ and $f_D(s)$. We first consider admission policies that take the form of a cutoff score (c_A and c_D) for each group so that students with scores above the cutoff for their group gain admission. We then consider more general admission policies.

We assume that the distribution of scores for D is worse, in a statistical sense, than the distribution for A. In particular, we assume that the likelihood ratio $\frac{f_D(s)}{f_A(s)}$ is weakly decreasing in s. The monotone likelihood ratio property implies that the the distribution $F_A(s)$ first-order stochastically dominates the distribution $F_D(s)$ ($F_A(s) \leq F_D(s)\ \forall s$). Under first-order stochastic dominance, any rational decision maker who prefers higher scores would choose a student from A over D if they knew only the student's group. It also implies that the mean score for D is lower than A ($\mu_D < \mu_A$). This property motivates our definition of a negative stereotype.

Definition 1 *A group D experiences a negative stereotype in relation to A if $F_A(s) \leq F_D(s)$ $\forall s$.*

Group-blind admissions will not eliminate the negative stereotype experienced by D. Suppose that the school wishes to accept students so that it achieves the highest average test score subject to the constraint that it can only accept a fixed proportion (K) of applicants. The admission problem can be formulated as:

$$\max_{c_D, c_A} \frac{1}{K} \cdot [\theta_D \int_{c_D}^{\overline{s}} s f_D(s) ds + \theta_A \int_{c_A}^{\overline{s}} s f_A(s) ds] \text{ subject to}$$
$$\theta_D \cdot (1 - F_D(c_D)) + \theta_A \cdot (1 - F_A(c_A)) = K$$

The optimal admission policy that solves this problem is group blind. It takes the form of a common cutoff score $c_D = c_A = c^*$ where c^* is determined by:

$$\theta_D \cdot (1 - F_D(c^*)) + \theta_A \cdot (1 - F_A(c^*)) = K$$

This policy does not eliminate the negative stereotype against admitted students from D. The likelihood ratio among admitted students is:

$$\frac{\frac{f_D(s)}{1-F_D(c^*)}}{\frac{f_A(s)}{1-F_A(c^*)}} = \frac{f_D(s)}{f_A(s)} \cdot \frac{1 - F_A(c^*)}{1 - F_D(c^*)}$$

By assumption, this ratio is weakly decreasing in s. Therefore, the distribution of scores for admitted students from A continues to first-order stochastically dominate the distribution for admitted students from D. The mean score of admitted students from A is also higher than the mean score of admitted students from D.

In fact, admitted students from D will experience a negative stereotype under *any* group-blind admission policy. Suppose that an admission policy consists of (i) a set of student characteristics, (ii) a score for each student based on these characteristics, and (iii) a probability of admission for each student based on this score. Such a policy is group blind if the student characteristics, score, and probability of admission do not vary by group.

Definition 2 *An admission policy is given by $[X, s(x), q_A(s), q_D(s)]$ where: (i) X denotes the set of student characteristics that are relevant for admission apart from membership in D or A, (ii) $s(x) : X \to \Re$ assigns a score to each student, and (iii) $q_A(s) : \Re \to [0, 1]$ and $q_D(s) : \Re \to [0, 1]$ assign a probability of admission for each score to types A and D.*

Definition 3 *An admission policy is group blind with respect to D and A if $q_A(s) = q_D(s)$ for all s.*

It follows that group-blind admission policies will not eliminate negative stereotypes.

Proposition 1 *If the likelihood ratio* $\frac{f_D(s)}{f_A(s)}$ *is weakly decreasing in* s *then, under any group-blind admission policy, admitted students from* D *experience a negative stereotype relative to admitted students from* A *.*

Proof. *The likelihood ratio among admitted students is given by:*

$$\frac{\frac{q_D(s)\cdot f_D(s)}{\int q_D(u)\cdot f_D(u)du}}{\frac{q_A(s)\cdot f_A(s)}{\int q_A(u)\cdot f_A(u)du}} = \left(\frac{\int q_A(u)\cdot f_A(u)du}{\int q_D(u)\cdot f_D(u)du}\right)\cdot\frac{f_D(s)}{f_A(s)}$$

because $\frac{q_D(s)}{q_A(s)} = 1$ *under group-blind admissions. By the monotone likelihood ratio property, this expression is strictly decreasing in* s*.* ■

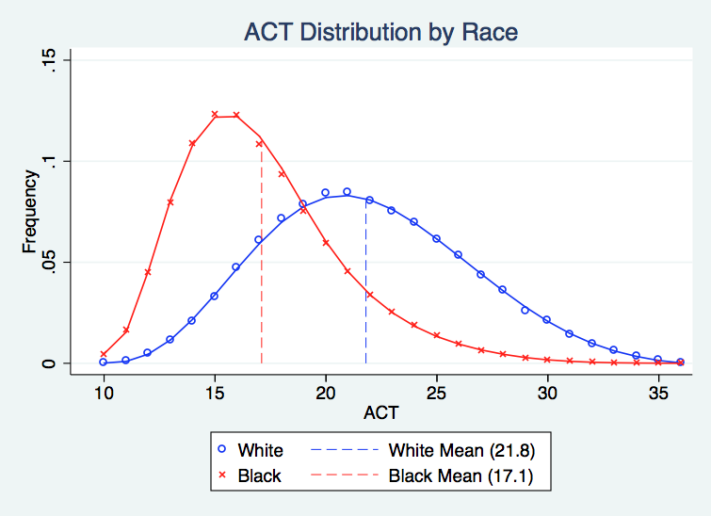
Under the monotone likelihood ratio assumption, group-blind admission policies will not eliminate negative stereotypes. If negative stereotypes are a cost of affirmative action, they are also a cost of group-blind admissions.

A.1 An ACT Illustration

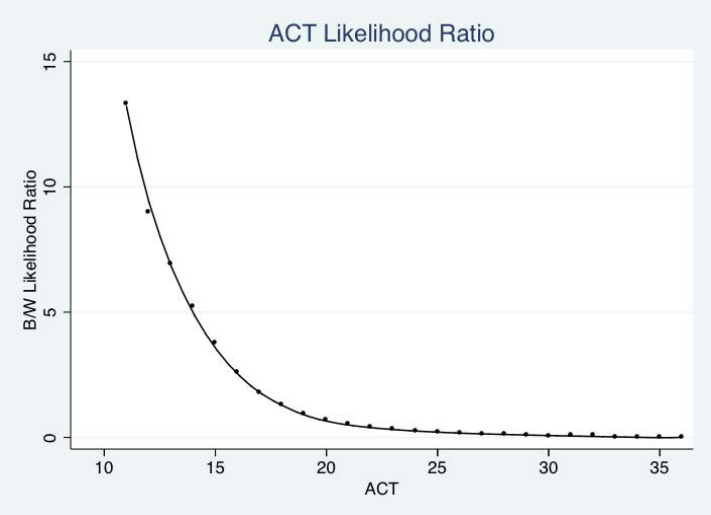
We illustrate the persistence of negative stereotypes under group-blind admissions using a random sample of ACT scores. We limit our data to students who self-identify as black or white on the ACT. We choose this example because the use of race in higher-education admissions is the most debated issue in affirmative action in the United States, and because tests like the ACT are an important, though by no means exclusive, component of admissions (Krueger, Rothstein, and Turner (2006)).[7] We take no position on whether, as an empirical matter, (i) such stereotypes are widely held or acted upon, or (ii) admitted students are harmed by it. Our aim is to demonstrate the persistence of a racial stereotype, as we have defined it, under a simple model of race-blind admissions.

The unconditional distributions of scores for both groups is depicted below.

[7] Internationally, it is not uncommon for admission to academic programs to be based on scores from a single test. For example, admission to the National Law Schools in India is based exclusively on the Common Law Admission Test. These law schools practice affirmative action on the basis of gender and caste.

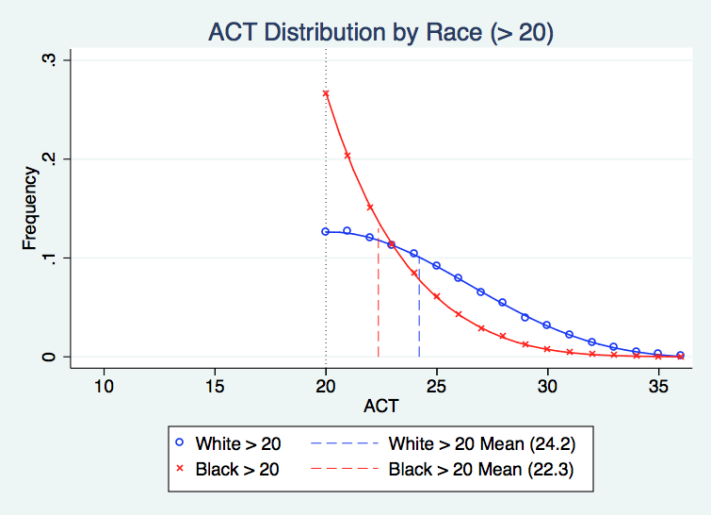


In our data, black students comprise 17% of test takers and white students comprise the remaining 83%. White students have a mean score of 21.8, and black students have a mean score of 17.1, implying a difference in means of 4.7. In terms of percentiles the mean black score is at the 18th percentile of the white distribution, while the mean white score is at the 92nd percentile of the black distribution. As seen in figure below, the two distributions satisfy the monotone likelihood ratio property. Therefore, the assumptions under which a negative stereotype exists under any race-blind admission policy are satisfied for this data. We suspect they are likely to hold in other cases where affirmative action policies are used.



We consider a race-blind admission policy in which the school admits the top 60% of students on the ACT. This implies admitting all students with ACT scores greater than or equal to 20. This policy maximizes the expected ACT score of admittees subject to

the constraint of admitting 60% of applicants. Under this policy, admitted white students continue to have a higher average ACT score than black students. This difference in means is 1.9, which is about 40% of the pre-existing difference (4.7). In terms of percentiles, the mean score for admitted black students is at the 37th percentile of the distribution for admitted white students. The mean for admitted white students is at the 88th percentile for admitted black students. A race-blind policy substantially reduces the percentage of black admitted students relative to the population of test takers. Black students would make up only 5.3% of admitted students, while white students would comprise the remaining 94.6%.



Black students would continue to suffer from a negative stereotype, even though admissions are race blind. The distribution of ACT scores for white admitted students first-order stochastically dominates the distribution for black students. Visually this can be seen from the fact that the likelihood ratio (D/A) among admitted students is decreasing. Affirmative action policies may leave in place negative stereotypes, but ending affirmative action would not eliminate these stereotypes. Therefore, critics of affirmative action who emphasize the harm of such stereotypes should be also be concerned with their effects under group-blind admissions.

B Eliminating Negative Stereotypes

If negative stereotypes are harmful to disadvantaged students, what would it take to eliminate them? We show that eliminating such stereotypes would require making admission standards *higher* for disadvantaged students than for advantaged students. This, in turn, would reduce the representation of disadadvantaged students among admittes to an even greater extent than a group-blind policy.

B.1 Equalizing Mean Scores

Consider again a simple admission policy that takes the form of a cutoff score and maximizes the expected score of admitted students. A policy that combats stereotypes by equalizing the mean score of admitted students from D and A solves:

$$\max_{c_D, c_A} \frac{1}{K} \cdot [\theta_D \int_{c_D}^{\overline{s}} s f_D(s) ds + \theta_A \int_{c_A}^{\overline{s}} s f_A(s) ds] \text{ subject to}$$

$$\theta_D \cdot (1 - F_D(c_D)) + \theta_A \cdot (1 - F_A(c_A)) = K \text{ and} \quad \text{(i)}$$

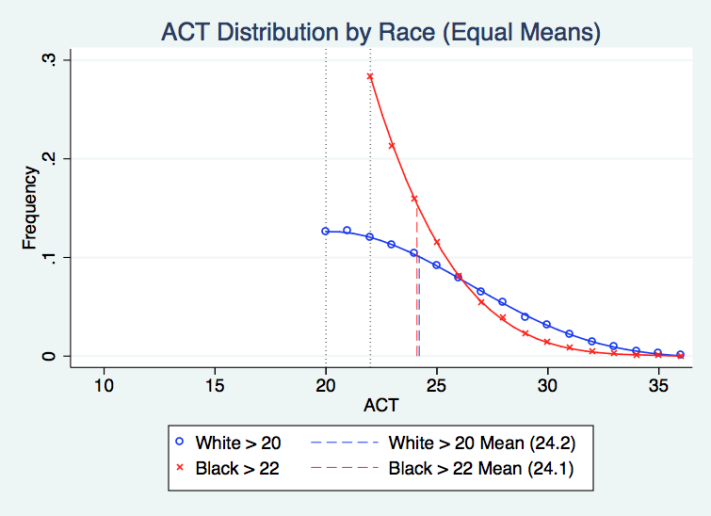
$$E_D[s|s > c_D] = E_A[s|s > c_A] \quad \text{(ii)}$$

The optimal policy (c_D^*, c_A^*) is determined entirely by the two constraints. From our previous analysis, we know that if $c_D^* \leq c_A^*$ admitted students from A have a higher mean score than admittees from D. It follows that $c_D^* > c_A^*$. Disadvantaged students must clear a *higher* threshold in order to gain admission.

There is a fundamental tension between a commitment to group-blind admissions and a commitment to eliminating stereotypes. Perversely, eliminating the negative stereotype against admitted students from D requires not equality, but making it more difficult for students from D to gain admission. This reduces the representation of admitted students from D even more than a group-blind policy. If negative stereotypes are indeed harmful, then an "equal-mean" policy would help admitted students from D, but it would do so by preventing other disadvantaged students from gaining admission.[8]

We illustrate this conclusion using data from the ACT. Below, we depict the optimal, (approximately) equal-mean policy $\{c_B^*, c_W^*\}$ that admits 60% of students.

[8] An equal-mean policy would also result in a lower expected score for admitted students, relative to a group-blind policy.



Black students are required to score above 22 to gain admission whereas white students are required to score 20. Under this equal-mean policy, black students comprise only 2.8% of admitted students. This is a 54% reduction in the representation of black students relative to a race-blind admission policy, for which black students comprise 5.3% of admitted students. This example illustrates that trying to eliminate stereotypes could dramatically and adversely affect the admission prospects of disadvantaged students.

B.2 Equalizing Distributions

Even if an admission policy equalizes the mean scores across admitted students from A and D, there will still be a larger fraction of students from A with very high scores. This can be seen by examining the distributions in our ACT example. Admission policies that assign a probability of admission to a score can completely eliminate stereotypes by equalizing the distribution of scores across A and D. However, eliminating stereotypes still requires making it more difficult for many disadvantaged students to gain admission.

Consider a general admission policy (Definition 2) that maximizes the expected score of admitted students, subject to the constraints that a constant proportion K of students are admitted and that the distribution of scores across admitted students from D and A are identical. Let ϕ_D and ϕ_A denote the probability of admission for each group, so that:

$$\phi_D = \int_s q_D(s) f_D(s) ds, \;\; \phi_A = \int_s q_A(s) f_A(s) ds$$

An optimal admission policy solves:

$$\max_{\{q_D(s),q_A(s),\phi_D,\phi_A\}} \frac{1}{K} \cdot [\theta_D \int_s s \cdot q_D(s) f_D(s) ds + \theta_A \int_s s \cdot q_A(s) f_A(s) ds] \text{ subject to}$$

$$\phi_D \cdot \theta_D + \phi_A \cdot \theta_A = K, \tag{i}$$

$$\frac{q_D(s) f_D(s)}{\phi_D} = \frac{q_A(s) f_A(s)}{\phi_A} \ \forall s, \tag{ii}$$

$$\phi_D = \int_s q_D(s) f_D(s) ds, \ \phi_A = \int_s q_A(s) f_A(s) ds, \tag{iii}$$

$$0 \leq q_D(s) \leq 1, \ 0 \leq q_A(s) \leq 1 \ \forall s, \ \textit{and} \tag{iv}$$

$$0 < \phi_D \leq 1, \ 0 < \phi_A \leq 1 \tag{v}$$

In contrast to contractual screening, in which a principal wishes to maximize an objective function while separating types, this admission policy can be thought of as contractual "shrouding." The principal wishes to maximize an objective function while preventing third parties or agents themselves from making inferences about each other's type.

Proposition 2 *A solution* $(q_D^*(s), q_A^*(s), \phi_D^*, \phi_A^*)$ *to this program exists and takes the form:*

$$q_D^*(s) = \left\{ \begin{array}{c} 0 \ if \ \underline{s} \leq s < s_L^* \\ \frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D^*}{\phi_A^*} \ if \ s_L^* \leq s < s_H^* \\ 1 \ if \ s_H^* \leq s < \overline{s} \end{array} \right\}$$

$$q_A^*(s) = \left\{ \begin{array}{c} 0 \ if \ \underline{s} \leq s < s_L^* \\ 1 \ if \ s_L^* \leq s < s_H^* \\ \frac{f_D(s)}{f_A(s)} \cdot \frac{\phi_A^*}{\phi_D^*} \ if \ s_H^* \leq s < \overline{s} \end{array} \right\}$$

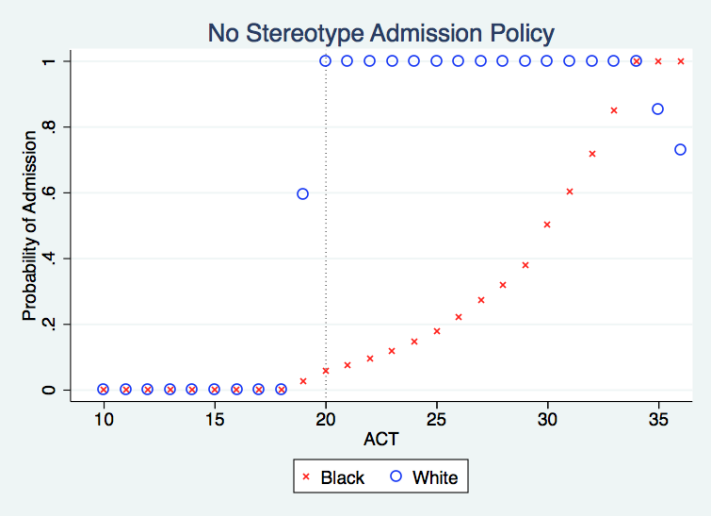
for appropriately chosen thresholds s_L^* *and* s_H^*.

Proof. *See Appendix* ■

Eliminating the negative stereotype requires making admission more difficult for applicants from D with scores in the interval $[s_L^*, s_H^*]$. The probability of admission for applicants from D is positive and increasing over this range, but it is less than the probability of admission for applicants from A. Applicants from D with scores in the upper range $[s_H^*, \overline{s}]$, however, are admitted with probability one and are more likely to gain admission than applicants from A. For applicants from A in this range, the probability of admission is less than 1 and is actually *decreasing*. Finally, because the threshold s_L^* is below the cutoff score under a group-blind policy, some students in both D and A are admitted under the

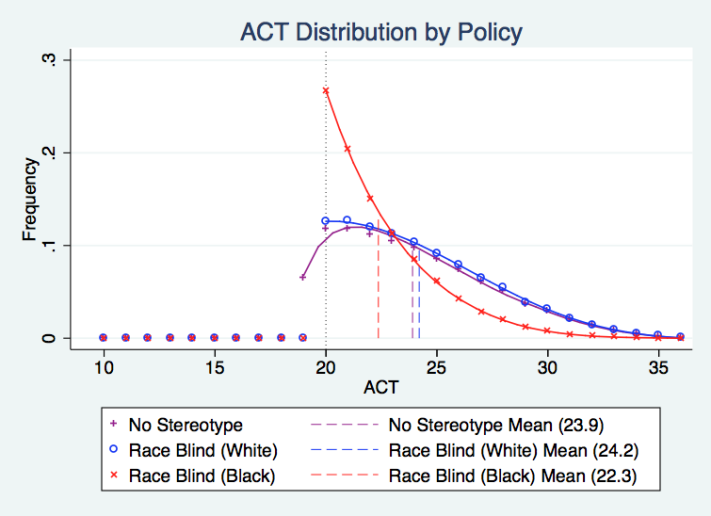
no-stereotype policy that would not gain admission under a group-blind policy.[9]

We again illustrate these conclusions using ACT data, where $q_B^*(s)$ and $q_W^*(s)$ represent the probability of admission for a black and white student, respectively, with score s.



Overall, it is more difficult for black applicants to gain admission than white applicants under a no-stereotype policy. Black students with SAT scores between 19 and 34 have a lower probability of admission than white students who score in this range. Black students who score 35 or more have a higher probability of admission than white students, but in total only 2.9% of black students are accepted, while 71.7% of white students are accepted. The percentage of black students among admitted students is .8%, while the percentage of white students is 99.2%. The no-stereotype distribution of scores among admitted students, which is identical for black and white admittees, is largely attained by matching the black distribution to the white distribution. This can seen by comparing the distribution of scores for black and white admittees under a no-stereotype policy to the distribution under a race-blind policy that admits all students who score over 20.

[9] If we constrain the probability of admission to be weakly increasing in s, then all students from A above a threshold s^* are admitted with probability 1 and students from D with a score s above this threshold are admitted with probability $q_D^*(s) = \frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D^*}{\phi_A^*}$ where $\phi_D^* = \frac{K}{\theta_A \frac{f_A(\overline{s})}{f_D(\overline{s})} + \theta_D}$ and $\phi_A^* = \frac{K}{\theta_A + \theta_D \frac{f_D(\overline{s})}{f_A(\overline{s})}}$. If $\frac{f_A(\overline{s})}{f_D(\overline{s})}$ is large, then the percentage of students from D is admitted is extremely small.



It is also more difficult for black students to gain admission under a no-stereotype policy than a race-blind policy. Black students who score between 20 and 33 on the ACT have a strictly lower probability of admission, while the probability is unchanged for those who score above 33. The percentage of black students among admittees (.8%) is lower than the percentage under a race-blind policy (5.3%) or a policy that equalizes the mean score of black and white admittees (2.8%).[10]

To summarize, negative stereotypes will continue to exist under group-blind admissions when admission scores satisfy the monotone likelihood ratio property, and this property holds in ACT data. If admission policies sought to eliminate group stereotypes, it would have dramatic consequences for students from disadvantaged groups. Disadvantaged students would be subject to even more stringent admissions standards than those applied to advantaged students. As a result, their representation among admitted students would fall substantially, even relative to group-blind admissions.

Admission policies are motivated by many values, but it is nevertheless useful to understand the consequences of particular values in isolation. Unlike a concern for stereotyping, other ethical values that inform higher-education admissions, when considered on their own, do not imply outcomes that very few persons would support. For example, a commitment to equal treatment implies that admission standards should be independent of group status. A commitment to equal representation may require admission quotas or other steps to match the characteristics of admitted students with those of the wider population. In comparison, the desire to eliminate stereotypes does not stand well on its own.

[10] Relative to a policy that equalizes the mean score of black and white admits, the probability of admission for black students under a no stereotype policy is higher for scores between 19 and 21, lower between 22 and 33, and equal when 34 and above.

II Other Values in Admissions

Eliminating stereotypes through admission policy would adversely affect disadvantaged students. We next consider how this conclusion changes when admission policies also embody other values. We have already shown there is a tension between a commitment to equal treatment and concern over negative stereotypes. Proposition 3 below states this precisely. When both these ends are valued, an optimal admission policy would still make it harder for disadvantaged students to gain admission. It requires a third value, such as a commitment to equal representation, for group-blind admission to be optimal. However, group-blind admissions are optimal only in the knife-edge case when the marginal gains from equal representation exactly offset the marginal cost of negative stereotypes. Some affirmative action is optimal whenever the former effect is larger.

A Stereotypes and Equal Treatment

We show that a commitment to both equal treatment and combating stereotypes is inconsistent with a group-blind admission policy. Suppose that admission policies again take the form of a cutoff score for each group. We assume that a school has a moral ordering over different policies that can be represented by the function:[11]

$$U(c_A - c_D,\ E_A[s|s > c_A] - E_D[s|s > c_D])$$

All else equal, a school prefers a policy with equal cutoffs c_A and c_D to one that differentiates applicants according to group status. The school also prefers an admission policy in which there is a smaller difference in the average scores of admitted students from each group. We assume that there are diminishing marginal returns to achieving the most desired outcome of equal cutoffs and equal means. In particular, we assume that (i) U is differentiable and attains its global maximum at $U(0,0)$, (ii) for any fixed x, $U(\cdot, x)$ achieves its maximum at $U(0,x)$ and $U(x,\cdot)$ achieves its maximum at $U(x,0)$, (iii) $U_1(0,x) = 0$ and $U_2(x,0) = 0$ for any x, and (iv) U is strictly concave. The admission problem is then:

[11] In our previous examples, a school maximizes the expected score of admitted students subject to different constraints. None of the following conclusions change if we also allow schools to value a higher expected score among admitted students. In this case, the utility function would include a third argument:

$$\frac{\theta_A(1 - F_A(c_A))}{\theta_A(1 - F_A(c_A)) + \theta_D(1 - F_D(c_D))} \cdot E_A[s|s > c_A] + \frac{\theta_D(1 - F_D(c_D))}{\theta_A(1 - F_A(c_A)) + \theta_D(1 - F_D(c_D))} \cdot E_D[s|s > c_D]$$

$$\max_{c_D, c_A} U(c_A - c_D,\ E_A[s|s > c_A] - E_D[s|s > c_D]) \text{ subject to}$$
$$\theta_D \cdot (1 - F_D(c_D)) + \theta_A \cdot (1 - F_A(c_A)) = K$$

We further assume that the distribution functions $F_A(s)$ and $F_D(s)$ are smooth enough that we can use the implicit function theorem to find a function $c_A(c_D)$ that will locally satisfy the constraint. Under these assumptions, the optimal admission policy would, perversely, make it harder for disadvantaged students to gain admission, despite the preference for equal treatment.

Proposition 3 *For the above preferences, an optimal admission policy will have a higher admission cutoff score for disadvantaged students ($c_D^* > c_A^*$).*

Proof. *Suppose that $c_A = c_D = c^*$. Define $c_A(c_D)$ so that the constraint is satisfied in an open interval containing c^*. Observe that $c_A'(c_D) < 0$ because increasing the cutoff for D requires lowering it for A to satisfy the constraint. The sign of the derivative of U with respect to c_D is given by:*

$$\frac{dU(c_A - c_D,\ E_A[s|s > c_A] - E_D[s|s > c_D)}{dc_D}|_{c_A = c_D = c^*} =$$
$$\underbrace{U_1(0,+)}_{0}\underbrace{\left(c_A'(c_D) - 1\right)}_{(-)} + \underbrace{U_2(0,+)}_{(-)}\underbrace{\frac{d(E_A[s|s > c_A(c_D)] - E_D[s|s > c_D])}{dc_D}|_{c_A = c_D = c^*}}_{(-)} > 0$$

Observe that (i) $U_1(0,+) = 0$ because $U(0,x)$ is a maximum and $U_1(0,x) = 0$, (ii) $U_2(0,+) < 0$ because $U(0,0)$ is a global maximum and U is strictly concave, and (iii) $\frac{d(E_A[s|s>c_A(c_D)] - E_D[s|s>c_D])}{dc_D}|_{c_A = c_D = c^*} < 0$ because raising c_D and lowering c_A lowers the mean difference in scores. Therefore $c_D^* = c_A^*$ cannot be an optimum because increasing c_D, while satisfying the constraint, increases the objective function. A similar argument shows that $c_D^* \not< c_A^*$. The sign of the derivative of U with respect to c_D is given by:

$$\frac{dU(c_A - c_D,\ E_A[s|s > c_A] - E_D[s|s > c_D)}{dc_D}|_{c_D^* < c_A^*} =$$
$$\underbrace{U_1(+,+)}_{(-)}\underbrace{\left(c_A'(c_D) - 1\right)}_{(-)} + \underbrace{U_2(+,+)}_{(-)}\underbrace{\frac{d(E_A[s|s > c_A(c_D)] - E_D[s|s > c_D])}{dc_D}|_{c_D^* < c_A^*}}_{(-)} > 0$$

Observe that (i) $U_1(+,+) < 0$ because $U(0,x)$ is a maximum and U is strictly concave, and (ii) $U_2(+,+) < 0$ because $U(x,0)$ is a maximum and U is strictly concave. Therefore $c_D^* < c_A^*$ cannot be an optimum. ■

Starting from group-blind admission policy ($c_D = c_A$), increasing the cutoff for students from D and correspondingly lowering it for students from A has two distinct effects. First, there is a loss from violating equal treatment, but this loss is small starting from an optimum. Second, there is a gain from lowering the mean difference in scores between the two groups and thereby reducing the stereotype against admitted students from D. This second effect dominates, so the school can do strictly better by increasing c_D. On the other hand, starting from an admission policy with affirmative action, where $c_D < c_A$, increasing the cutoff for students from D and correspondingly lowering it for students from A has two salutary effects: first, it decreases inequality of treatment, and second it decreases the loss from stereotyping.

The optimal policy would not eliminate stereotyping. At that point, reducing the inequality of cutoffs by a small amount would have first-order gains from increased equality of treatment and only second-order losses from increased stereotyping. Admissions cutoffs would be higher for disadvantaged applicants, but not so high as to eliminate stereotyping. An optimal admission policy in this setting would entail both unequal treatment and stereotyping.

These conclusions are closely related to those of the previous section. Admitted students from D experience a negative stereotype under group-blind admissions. Therefore, a school that values both equal treatment and combating stereotypes would increase admission standards for disadvantaged groups in order to limit stereotyping.[12]

A.1 Other Preferences Over Equal Treatment and Stereotypes

There are other ways to describe preferences over equal treatment and stereotypes that do not imply this result. In our view, the best way to represent the preferences of advocates of group-blind admissions is to accord a lexicographic priority to equal treatment. Admission policies would first be ranked in terms of their proximity to a group-blind policy. If there are any ties on this dimension, they would then be ranked in terms of their effect on negative stereotypes. Such preferences reflect a concern for *both* equal treatment and combating stereotypes, but the optimal admission policy under these preferences would be group blind. Because preferences are revealed through choices, a lexicographic preference for equal treatment will usually be *observationally equivalent* to preferences that completely ignore stereotypes. Therefore combating negative stereotypes should not play an important

[12] Note that advocating group-blind admissions is distinct from arguing that, starting from affirmative action, admission policy should move in the direction of equal treatment. The latter argument is consistent with the preferences we have described for equal treatment and reducing stigma, but the former is not. With these preferences, one should move not just toward equal treatment, but past it.

role when making the case for group-blind admissions.[13]

Are there other preferences over equal treatment and stereotypes that imply group-blind admissions? The choice of group-blind admissions is a "corner solution" in a decision problem where the other corner is eliminating stereotypes. We cannot analyze the entire set of preferences that would produce this outcome, but we can consider other plausible alternatives. For example, corner solutions often arise for linear preferences where the marginal benefit of increasing equality or reducing stereotypes is constant and does not depend on the level of equality or stereotype.

Linear preferences do not adequately explain group-blind admissions. Suppose an opportunity arises in which a small change in the equality of treatment can result in a large change in the stereotype experienced by the disadvantaged group. Lowering the standard for the advantaged group could result in admitting a large number of students that would lower the average score for that group. Under linear preferences, a decision maker may then switch to a no-stereotype policy. This logic also applies if there are increasing marginal returns to equalizing treatment and reducing stereotypes. Therefore, we do not believe that such preferences adequately rationalize the choice of group-blind admissions.

Nor are discontinuous preferences, which reflect a large, fixed cost of moving away from group-blind admissions, a plausible alternative. Consider a group of candidates (*tier* 1) that a school would strictly prefer to admit on the basis of their scores and a second group of candidates (*tier* 2) that the school is indifferent between admitting or not. Compare (i) the decision to favor candidates from D in *tier* 2 over candidates from A in *tier* 2 to (ii) the decision to favor a candidate in D from *tier* 2 over a candidate in A from *tier* 1. If there is a large fixed cost to violating equal treatment, then a decision maker would perceive a greater ethical violation in moving from a group-blind policy to (i), than in moving from (i) to (ii). We view this as implausible. Moreover, the existence of a large fixed cost of moving away from group-blind admissions would be observationally equivalent to advocating group-blind admissions in all circumstances.

We conclude that a lexicographic preference for equal treatment best explain the support for group-blind admissions. This conclusion also applies to the *constitutional* view articulated by Justice Thomas. In his concurring opinion in *Fisher v. Texas*, Justice Thomas states that "for constitutional purposes" it does not matter whether the effects of affirmative action are "insidious" or "benign." In our view, Justice Thomas's statement also describes normative arguments for group-blind admissions that do not explicitly rely

[13] Group-blind admissions would also be optimal if a school had a constraint (or lexicographic preference) that disadvantaged groups be treated no worse than advantaged groups in the admission process. This can be described as an "antisubordination principle." If a school had, in addition, preferences over equal treatment and combating stereotypes with diminishing margins, the antisubordination principle would result in the choice of group-blind admission. We thank David Weisbach for bringing this point to our attention.

on constitutional interpretation.

B Stereotypes, Equal Treatment, and Equal Representation

Finally we consider admissions when equal representation is valued in addition to equal treatment and combating stereotypes. We define equal representation as a desire for the demographics of admitted students to be similar to the general population. Equal representation can also be described as a concern for diversity. Other goals of affirmative action policies, such as remedying past or current discrimination, will usually be satisfied by moves toward more equal representation.

We show that it requires a third value such as equal representation for group-blind admissions to be optimal. However, group blind admissions are only optimal in the knife-edge case when preferences over equal representation and stereotypes exactly balance. We suggest that this perfect balance does not adequately capture the case for group-blind admission. When all three of these ends are valued, the most plausible conclusion is that some amount of affirmative action is optimal.

Consider a school with preferences over equal treatment and combating stereotpes. The school also prefers more equal representation.[14] All else equal, the school would like the relative proportion of admitted students from groups A and D be as close as possible to their population proportion. These preferences can be represented by the function:[15]

$$U(c_A - c_D, E_A[s|s > c_A] - E_D[s|s > c_D], F_D(c_D) - F_A(c_A))$$

We again assume that there are diminishing marginal returns to achieving the most desired outcome of equal cutoffs, equal means, and equal representation. In particular, we assume that (i) U attains its global maximum at $U(0,0,0)$, (ii) $U(0,x,y)$, $U(x,0,y)$, and $U(x,y,0)$ are maxima for any fixed x and y, (iii) $U_1(0,x,y) = 0$, $U_2(x,0,y) = 0$, $U_3(x,y,0) = 0$ for any x and y, and (iv) U is strictly concave. The admissions problem is then:

[14] As in the previous section, our conclusions do not change if we also allow the school to prefer a higher expected score among admitted students.

[15] If the ratio of advantaged to disadvantaged students in the population is $\frac{\theta_A}{\theta_D}$, then the ratio among admitted students is:

$$\frac{\frac{\theta_A(1-F_A(c_A))}{\theta_A(1-F_A(c_A)+\theta_D(1-F_D(c_D))}}{\frac{\theta_D(1-F_D(c_D))}{\theta_A(1-F_A(c_A))+\theta_D(1-F_D(c_D))}} = \frac{\theta_A}{\theta_D} \cdot \frac{(1-F_A(c_A))}{(1-F_D(c_D))}$$

and the difference between the two ratios is

$$\frac{\theta_A}{\theta_D} \cdot (\frac{(1-F_A(c_A))}{(1-F_D(c_D))} - 1)$$

This difference is zero if and only if the probability of acceptance for both groups is equal.

$$\max_{c_D,c_A} U(c_A - c_D,\ E_A[s|s > c_A] - E_D[s|s > c_D], F_D(c_D) - F_A(c_A)) \text{ subject to}$$
$$\theta_D \cdot (1 - F_D(c_D)) + \ \theta_A \cdot (1 - F_A(c_A)) = K$$

Proposition 4 *For these preferences, group-blind admissions are optimal only if the marginal benefit of reducing stereotype equals the marginal cost of lower representation.*

Proof. *Suppose that the group-blind policy ($c_A = c_D = c^*$) that satisfies the constraint is optimal. Consider effect of increasing c_D while continuing to satisfy the constraint. Because c^* is optimal:*

$$\frac{dU(c_A - c_D,\ E_A[s|s > c_A] - E_D[s|s > c_D], F_D(c_D) - F_A(c_A))}{dc_D}|_{c_A=c_D=c^*} =$$
$$\underbrace{U_2(0,+,+) \cdot \frac{d(E_A[s|s > c_A(c_D)] - E_D[s|s > c_D])}{dc_D}|_{c_A=c_D=c^*}}_{MB(+)} +$$
$$\underbrace{U_3(0,+,+) \cdot \frac{d(F_D(c_D) - F_A(c_A(c_D)))}{dc_D}|_{c_A=c_D=c^*}}_{MC(-)} = 0$$

We know that (i) $U_2(0,+,+) < 0$ because $U(x,0,y)$ is a maximum and U is strictly concave, (ii) $\frac{d(E_A[s|s>c_A(c_D)]-E_D[s|s>c_D])}{dc_D}|_{c_A=c_D=c^*} < 0$ from Proposition 3, (iii) $U_3(0,+,+) < 0$ because $U(x,y,0)$ is a maximum and U is strictly concave, and (iv) $\frac{d(F_D(c_D)-F_A(c_A(c_D)))}{dc_D}|_{c_A=c_D=c^*} > 0$ because raising c_D and lowering c_A increases this difference. Therefore c^* is an optimum only if the marginal benefit of reducing the mean difference in scores exactly equals the marginal cost of increasing the difference in representation. ■

From Proposition 3, we know that starting from group-blind admissions, the marginal cost of violating equal treatment by increasing c_D (and lowering c_A) is small. Therefore, for a group blind policy to be optimal the gain from lowering the average difference in test scores must exactly equal the loss from reducing the representation of disadvantaged students.

When there are diminishing margins, it requires a third value–such as equal representation– to rationalize a policy of group-blind admissions. Nevertheless, we find it unlikely that such a delicate balancing act explains the advocacy of group-blind admissions. First, few critics of affirmative action who emphasize the harm to recipients from stereotypes also voice an equal concern for promoting equal representation through admission policy. Second, if many schools found these tradeoffs to perfectly balance, then given any heterogeneity among schools, we should expect some to favor further reducing stereotypes and to actually

impose a higher admissions cutoff on disadvantaged students. Yet no one takes seriously the position that admission standards should be higher for disadvantaged students.

Third, the balance between equal treatment and stereotypes should vary with other circumstances, such as a school's quota constraint. For example, suppose group-blind admissions are optimal for a large, less selective school (K_L) that admits all students above c_L. If the schools have similar preferences over equal treatment and stereotypes, then the choice of group-blind admissions at the large, less selective school should inform the choice of admission policy for a small, highly selective school ($K_S < K_L$). For the highly selective school, under group-blind admissions the mean difference in scores between admitted students from A and D will be lower[16], but the ratio of A to D among admittees will be higher.[17] Relative to the less selective school, the marginal value of increasing $D's$ representation is higher and the marginal value of reducing $D's$ stereotype is lower. Therefore, some amount of affirmative action should be chosen at the more selective school. By the same logic, if group-blind admissions are optimal at the more selective school, then the admission standards should be higher for students from D at the less selective school. Yet no advocates of group-blind admissions would vary their prescriptions in this way.

We conclude that it is unlikely that group-blind admissions are explained by equally-balanced preferences over equal representation and stereotypes. In contrast, some affirmative action is optimal whenever, starting from group-blind admissions, the marginal benefit of more equal representation outweighs the marginal cost of greater stereotypes. This will almost always be the case. Again, if it were not then some schools should advocate higher admission standards for disdvantaged students. Many would agree that equal treatment, equal representation, and combating stereotypes are worthy goals of admission policy. But no one would advocate such a perverse double-standard.

[16] This follows for some K_S in our model because:

$$\lim_{c \to \bar{s}} \{E[s|s \geq c] - c\} = 0$$

Even if s were unbounded, it would be true for some K_S as long as:

$$\lim_{c \to \infty} \{E[s|s \geq c] - c\} = 0$$

This is true, for example, for a normal distribution.

[17] For any group-blind admission policy with cutoff score c, the ratio of A to D among admitted students is:

$$\frac{\theta_A}{\theta_D} \cdot \frac{(1 - F_A(c))}{(1 - F_D(c))}$$

The monotone likelihood ratio property implies that this ratio is increasing in c.

III Other Views of Stereotypes

We focus on statistical discrimination to investigate the claim that stereotypes harm beneficiaries of affirmative action. Models of statistical discrimination have been used since at least Arrow (1972, 1973) and Phelps (1972). Our equation of stereotype with statistical discrimination is consistent with how test scores and other measures of academic performance are used in many discussions of affirmative action. For example, the admission policy at issue in *Fisher v. Texas* is determined on the basis of a numerical Academic Index (AI) and a Personal Achievement Index (PAI). The AI is calculated from a student's standardized test scores, class rank, and high school coursework. The PAI is based on a holistic review of the application and can take race into account. In our model, some minority groups would experience a negative stereotype with respect to their AI score or its components. In his concurring opinion, Justice Thomas points out that black and hispanic students admitted outside of the University of Texas's Top Ten Percent Plan have lower average GPAs and SAT scores than their white or asian counterparts. Consistent with our results, however, these differences are substantially similar for candidates admitted *under* Texas's Top Ten Percent program.[18] When schools adopt race-blind admission criteria that are intended to achieve diversity, stereotypes in the form of average test-score differences will likely still exist.

For these figures, Justice Thomas relied on an amicus brief filed by Richard Sander, a law professor at UCLA and attorney Stuart Taylor. Sander is well known for his criticism of affirmative action programs in law schools (Sander (2004)). However, the use of differences in test scores has also been used to measure the extent of affirmative action and "stereotype threat" by scholars who are broadly supportive of affirmative action. For example, in their widely-cited study, Fischer and Massey (2007) measure "affirmative action at the institutional level by taking the difference between the average SAT score earned by blacks or Hispanics and all students at a particular institution." The authors use this measure to test for the effect of stereotypes on affirmative action recipients.

Finally, the logic of our argument is consistent with any "stereotype" (however defined) that (i) continues to exist under group-blind admissions, and (ii) decreases if admissions are made more difficult for the disadvantaged group. This includes stereotypes based on false priors about the academic performance of disadvantaged groups. It includes some accounts of "external stigma"–the perception of affirmative action recipients by others–in social psychology (Bowen (2010)). It can also include accounts of "internal stigma" such

[18] According to the University of Texas at Austin's Office of Admissions, the average SAT/ACT scores of entering freshmen admitted under the top 10% program in 2009 by ethnic group were: White (1864/28), African-American (1584/22), Asian-American (1874/28), and Hispanic (1628/23). For entering freshmen admitted outside the program, the figures are: White (1914/29), African-American (1524/20), Asian-American (1991/30), and Hispanic (1794/27).

as "stereotype threat" (Steele and Aronson (1995), (Steele (1997)) if these effects decrease when admission standards are raised for disadvantaged groups.

Our conclusions do not follow for stereotypes that *only exist* in the presence of affirmative action. Suppose that a negative stereotype is defined as the probability that a student is admitted because of affirmative action. This is a reasonable interpretation of Justice Thomas's statements on the negative effect of stereotypes. This probability would increase with the extent of affirmative action, but ending affirmative action would eliminate the stereotype. This definition can be thought of as statistical discrimination, but only on the extensive margin of admissions. While plausible, this definition is incomplete. Admission under a competitive process is valuable signal only if it provides information concerning some underlying attribute. Therefore, statistical discrimination should occur on the basis of this attribute in addition to the fact of admission itself.

Ultimately, our claim is not that every form of stereotype or discrimination against affirmative action recipients would continue to exist under group-blind admissions. For example, there is evidence that negative inferences against affirmative-action recipients are mediated by the perceived fairness of the admissions procedure (Heilman, McCullough, and Gilbert(1996), Evans (2003)). Our claim is that many of them would.

IV Conclusion

In this paper, we analyze the use of admission policy to affect stereotypes against students from disadvantaged groups. Many critics of affirmative action in admissions argue that lower standards cause such stereotypes. We show that when stereotypes are a result of social inequality, they can easily persist under group-blind admissions. Perversely, eliminating stereotypes requires making it harder for disadvantaged students to gain admission. Such an admission policy would increase social inequality and is clearly unacceptable.

This conclusion does not change if schools seek to both treat students equally and counteract stereotypes. Under these preferences, an optimal admission policy would still impose a higher standard on disadvantaged students. Group-blind admissions are optimal, however, if a school holds an ethically prior, lexicographic preference for equal treatment. We argue that the best case for group-blind admissions rests on such a preference. Consequently, stereotypes should play a negligible role in critiques of affirmative action that advocate group-blind admissions.

If a school is willing to consider tradeoffs between equal treatment and stereotypes, it requires a third goal such as equal representation to justify group-blind admissions. However, group-blind admissions are optimal only when the conflicting goals of equal representation and limiting stereotypes exactly balance. In our view, this knife edge case is an implausible

justification for group-blind admission. It would again imply a desire on the part of some schools to have higher standards for disadvantaged students. Consequently, some amount of affirmative action is almost always optimal if a school values all three of these goals.

References

[1] Arrow, Kenneth, "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973, pp. 3-33.

[2] Becker, Gary S., *The Economics of Discrimination*, Chicago: University of Chicago Proess, 1957.

[3] Bowen, Deirdre M. "Brilliant Disguise: An Empirical Analysis of a Social Experiment Banning Affirmative Action," 85 *Indiana Law Journal* 1197 (2010).

[4] Bowen, William G., and Derek Bok, *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*, Princeton: Princeton University Press, 1998.

[5] Chan, Jimmy and Erik Eyster, "Does Banning Affirmative Action Lower College Student Quality?" *American Economic Review*, Vol. 93, No. 3 (2003) pp. 858-872.

[6] Chan, Jimmy and Erik Eyster, "The Distributional Consequences of Diversity-Enhancing University Admission Rules," *Journal of Law, Economics, and Organization* 25(2) (2009) pp. 499-517.

[7] Coate, Stephen, and Glenn C. Loury, "Will Affirmative-Action Politicies Eliminate Negative Stereotypes?" *The American Economic Review*, Vol. 83, No. 5 (Dec., 1993) pp. 1220-1240.

[8] Evans, David C., "A Comparison of the Other-Directed Stigmatization Produced by Legal and Illegal Forms of Affirmative Action," *Journal of Applied Psychology*, 88(1) pp. 121-30 (2003).

[9] Fang, Hanming, and Andre Moro,"Theories of Statistical Discrimination and Affirmative Action: A Survey," in Benhabib, J., Bisin, A., Jackson, M., (Eds.), *Handbook of Social Economics*: Elsevier Publishing, Amsterdam, pp. 133-200 (2010).

[10] Fischer, Mary J. and Douglas S. Massey, "The effects of affirmative action in higher education," *Social Science Research*, Volume 36, Issue 2, pp. 531-549 (2007).

[11] Fryer, Roland, Glenn Loury and Tolga Yuret, "An Economic Analysis of Color-Blind Affirmative Action," *Journal of Law, Economics, and Organization*, 24(2) pp. 319-335 (2008).

[12] Fryer, Roland, "Belief Flipping in a Dynamic Model of Statistical Discrimination," Journal of Public Economics, 91(5-6) pp. 1151-1166 (2007).

[13] Heilman, M.E., W.F. McCullough, and D. Gilbert, "The Other Side of Affirmative Action: Reactions of Non-Beneficiaries to Sex-Based Preferential Selection," *Journal of Applied Psychology*, 81 pp. 346-57 (1996).

[14] Kruger, Alan, Jesse Rothstein, and Sarah Turner, "Race, Income, and College in 25 Years: Evaluating Justice O'Connor's Conjecture," *American Law and Economics Review*, 8 pp. 282-311 (2006).

[15] Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, September 1972, 62, 659-61.

[16] Ray, Debraj and Rajiv Sethi, "A Remark on Color-Blind Affirmative Action," *Journal of Public Economic Theory*, 12(3) pp. 399-406 (2010).

[17] Sander, Richard, "A Systematic Analysis of Affirmative Action in American Law Schools," *Stanford Law Review* 57, pp 368-478 (2004).

[18] Steele, C.M., and J. Aronson, "Stereotype Threat and the Intellectual Test-Performance of African Americans," *Journal of Personality and Social Psychology*, 69(5) pp. 797-811 (1995).

[19] Steele, C.M., "How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, 52(6) pp. 613-629 (1997).

V Appendix

Proof of Proposition 2

A no-stereotype admission policy is a solution to the program:

$$\max_{\{q_D(s),q_A(s)\}} \frac{1}{K} \cdot \left[\theta_D \int_s s \cdot q_D(s) f_D(s) ds + \theta_A \int_s s \cdot q_A(s) f_A(s) ds\right] \text{ subject to}$$

$$\phi_D \cdot \theta_D + \phi_A \cdot \theta_A = K, \tag{i}$$

$$\frac{q_D(s) f_D(s)}{\phi_D} = \frac{q_A(s) f_A(s)}{\phi_A} \ \forall s, \tag{ii}$$

$$\phi_D = \int_s q_D(s) f_D(s) ds, \ \phi_A = \int_s q_A(s) f_A(s) ds, \tag{iii}$$

$$0 \leq q_D(s) \leq 1, \ 0 \leq q_A(s) \leq 1 \ \forall s, \text{ and} \tag{iv}$$

$$0 < \phi_D \leq 1, \ 0 < \phi_A \leq 1 \tag{v}$$

We restrict our attention to the case where $\frac{K}{\theta_D} < 1$ and $\frac{K}{\theta_A} < 1$. This ensures that the constraint set is nonempty. In particular, for any choice of ϕ_D and ϕ_A that satisfy constraints (i) and (v), there exist $q_D(s)$ and $q_A(s)$ that satisfy constraints (ii)-(iv).[19] We first prove the existence of a solution $(q_D^*(s), q_A^*(s), \phi_D^*, \phi_A^*)$ to this problem and then further characterize the solution. We can substitute for $q_A(s)$ using constraint (ii) and rewrite the program as:

$$\max_{\{q_D(s),\phi_D,\phi_A\}} \frac{1}{\phi_D} \cdot \int_s s \cdot q_D(s) f_D(s) ds \text{ subject to}$$

$$\int_s q_D(s) f_D(s) ds = \phi_D \tag{1}$$

$$\phi_D \cdot \theta_D + \phi_A \cdot \theta_A = K \tag{2}$$

$$0 \leq q_D(s) \leq \min[\frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D}{\phi_A}, 1] \tag{3}$$

$$0 < \phi_D \leq 1, \ 0 < \phi_A \leq 1 \tag{4}$$

Lemma 1 *For any fixed $\phi_D \in (0, \frac{K}{\theta_D})$, $\phi_A(\phi_D) = \frac{K - \theta_D \phi_D}{\theta_A}$ that satisfy constraints (2) and*

[19] If it is possible to admit all members of one group ($\phi_D = 1$ or $\phi_A = 1$) and satisfy the quota constraint, then it may not be possible to equalize the distribution of scores for admitted students for some values of ϕ_D and ϕ_A.

(4), a solution to the program exists and takes the form:

$$q_D^*(s;\phi_D) = \left\{ \begin{array}{c} 0 \; if \; s < s_L^*(\phi_D) \\ \min[\frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D}{\phi_A(\phi_D)}, 1] \; if \; s \geq s_L^*(\phi_D) \end{array} \right\}$$

where $s_L^*(\phi_D)$ *is implicitly defined by:*

$$\int_{s_L^*(\phi_D)}^{\overline{s}} \min[1, \frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D}{\phi_A(\phi_D)}] f_D(s) ds = \phi_D$$

The solution function $q_D^*(s;\phi_D)$ *is weakly increasing in* s*, and the value function:*

$$\frac{1}{\phi_D} \cdot \int_{s_L^*(\phi_D)}^{\overline{s}} s \cdot q_D^*(s;\phi_D) f_D(s) ds$$

is continuous in ϕ_D*.*

Proof. For any choice of ϕ_D that satisfies constraints (2) and (4), the objective function is a continuous function defined over the set of functions $q(s)$ that satisfy constraints (1) and (3), and this set is compact. Therefore, a solution exists and by the maximum theorem, the value function is continuous in ϕ_D. Because the objective function is linear in $q(s)$, the maximum occurs at a "bang-bang" solution with respect to the constraint set. ■

Lemma 2 *A solution to this program* $(q_D^*(s), q_A^*(s), \phi_D^*, \phi_A^*)$ *exists where:*

$$\phi_D^* \in \left[\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})} \right) + \theta_D}, \frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})} \right) + \theta_D} \right]$$

Proof. We can express the "concentrated" program as:

$$\max_{\phi_D \in (0, \frac{K}{\theta_D})} \frac{1}{\phi_D} \cdot \int_{s_L(\phi_D)}^{\overline{s}} s \cdot q_D^*(s;\phi_D) f_D(s) ds$$

Consider any $\phi_D < \frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})} \right) + \theta_D}$. This implies $\frac{f_A(\overline{s})}{f_D(\overline{s})} \cdot \frac{\phi_D}{\phi_A(\phi_D)} < 1$ and that $q_D^*(s;\phi_D) = \frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D}{\phi_A(\phi_D)} < 1 \; \forall s \in [s_L^*(\phi_D), \overline{s}]$. The objective function is $E_A[s | s > s_L^*(\phi_D)]$ where $s_L^*(\phi_D) = 1 - F_A(\phi_A(\phi_D))$. Therefore, the objective function is strictly increasing in ϕ_D on $\left(0, \frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})} \right) + \theta_D} \right)$. Consider next any $\phi_D > \frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})} \right) + \theta_D}$. This implies $\frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D}{\phi_A} > 1 \; \forall s$ and $\phi_D = 1 - F_D(s_L(\phi_D))$. The objective function is $E_D[s | s > s_L(\phi_D)]$, which

is strictly decreasing in ϕ_D. We can therefore limit consideration to ϕ_D in the closed interval $\left[\frac{K}{\theta_A\cdot\left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right)+\theta_D}, \frac{K}{\theta_A\cdot\left(\frac{f_A(\underline{s})}{f_D(\underline{s})}\right)+\theta_D}\right]$. A maximum ϕ_D^* exists because we are evaluating a continuous function over a compact set. This implies the existence of ϕ_A^* from constraint (2) and $(q_D^*(s), q_A^*(s))$ from the previous Lemma. ■

Lemma 3 *For any solution* $(q_D^*(s), q_A^*(s), \phi_D^*, \phi_A^*)$, *(i)* $q_D^*(s) = 1$ *on* $[s_H^*, \overline{s}]$ *where* $s_L^* \leq s_H^* \leq \overline{s}$ *and (ii)* $\frac{f_A(s_H^*)}{f_D(s_H^*)} \cdot \frac{\phi_D^*}{\phi_A^*} = 1$.

Proof. We know $\frac{f_A(\overline{s})}{f_D(\overline{s})} \cdot \frac{\phi_D^*}{\phi_A^*} \geq 1$ because this inequality holds for all:

$$\phi_D \in \left[\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}, \frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})}\right) + \theta_D}\right]$$

Therefore $q_D^*(\overline{s}) = 1$, which proves (i). If $\frac{f_A(s_H^*)}{f_D(s_H^*)} \cdot \frac{\phi_D^*}{\phi_A^*} > 1$ this implies $s_L^* = s_H^* = s^*$. But then the objective function is $E_D[s|s > s^*]$, which is strictly decreasing in ϕ_D. This proves (ii). ■

The choice of an optimal, no-stereotype admission policy on the space of functions $(q_D(s), q_A(s))$ is therefore equivalent to a Kuhn-Tucker program in four, real-valued choice variables $(\phi_A, \phi_D, s_L, s_H)$:

$$\max_{\{\phi_A,\phi_D,s_L,s_H\}} \frac{1}{\phi_A} \cdot \int_{s_L}^{s_H} s \cdot f_A(s)ds \; + \frac{1}{\phi_D} \cdot \int_{s_H}^{\overline{s}} s \cdot f_D(s)ds \text{ subject to}$$

$$\phi_D \cdot \theta_D + \phi_A \cdot \theta_A = K \tag{a}$$

$$\frac{f_A(s_H)}{f_D(s_H)} \cdot \frac{\phi_D}{\phi_A} = 1 \tag{b}$$

$$\frac{F_A(s_H) - F_A(s_L)}{\phi_A} + \frac{1 - F_D(s_H)}{\phi_D} = 1 \tag{c}$$

$$\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D} \leq \phi_D \leq \frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})}\right) + \theta_D} \tag{d}$$

$$\frac{K}{\theta_A + \theta_D \cdot \left(\frac{f_D(\underline{s})}{f_A(\underline{s})}\right)} < \phi_A \leq \frac{K}{\theta_A + \theta_D \cdot \left(\frac{f_D(\overline{s})}{f_A(\overline{s})}\right)} \tag{e}$$

$$\underline{s} \leq s_L \leq s_H \leq \overline{s} \tag{f}$$

Lemma 4 *The solution function* $q_D^*(s) = 1$ *on some interval* $[s_H^*, \overline{s}]$ *where* $s_L^* \leq s_H^* < \overline{s}$.

Proof. Suppose that $\phi_D^* = \frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}$ so that $s_H^* = \overline{s}$ and $\frac{f_A(\overline{s})}{f_D(\overline{s})} \cdot \frac{\phi_D^*}{\phi_A^*} = 1$. This implies that all students from group A with scores above $s_L^* = F_A^{-1}\left(1 - \frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})}\right) + \theta_D}\right)$ are admitted with probability 1 ($q_A^*(s) = 1 \; \forall s > s_L^*$) and that students from D with scores in this range are admitted with probability weakly less than 1 ($q_D^*(s) = \frac{f_A(s)}{f_D(s)} \frac{f_D(\overline{s})}{f_A(\overline{s})} \leq 1$). To show that this cannot be the case, we consider the effect of slightly increasing ϕ_D when $\phi_D = \frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}$. We first describe the set of ϕ_D for which the equality constraints ((a)-(c)) all hold. Consider the implicit function $s_{\frac{A}{D}}(\phi_D)$ defined by constraints (a) and (b). This function is continuous and strictly decreasing on the interval defined in (d) and is differentiable everywhere on its interior. Further we know $s_{\frac{A}{D}}\left(\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}\right) = \overline{s}$ so that constraint (b) must hold and $s_{\frac{A}{D}}\left(\frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})}\right) + \theta_D}\right) = \underline{s}$ so that constraint (b) cannot hold. Consider next the implicit function $s_D(\phi_D)$ defined by $1 - F_D(s_H) = \phi_D$, which is also continuous and strictly decreasing on the interval defined in (d) and is differentiable everywhere on it interior. We know that $s_D\left(\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}\right) < \overline{s}$ and $s_D\left(\frac{K}{\theta_A \cdot \left(\frac{f_A(\underline{s})}{f_D(\underline{s})}\right) + \theta_D}\right) > \underline{s}$. By the intermediate value theorem, there exist a set of $\widehat{\phi}_D$ in this interval such that $s_{\frac{A}{D}}(\phi_D) = s_D(\phi_D)$. Because this set is compact, we can define $\underline{\widehat{\phi}}_D$ to be the minimum over this set so that for all $\phi_D \in \left[\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}, \underline{\widehat{\phi}}_D\right]$, constraints (a)-(c) hold with equality. Therefore, for any choice of ϕ_D in this interval, the equality constraints define continuous, implicit functions $\phi_A(\phi_D), s_L(\phi_D)$,and $s_H(\phi_D)$. By the implicit function theorem, these functions are differentiable in the open interval $\left(\frac{K}{\theta_A \cdot \left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right) + \theta_D}, \underline{\widehat{\phi}}_D\right)$ with derivatives given by:

$$\phi_A^{'}(\phi_D) = -\frac{\theta_D}{\theta_A}$$

$$s_L^{'}(\phi_D) = \frac{\phi_A}{f_A(s_L)} \cdot \left[\frac{F_A(s_H) - F_A(s_L)}{\phi_A^2} \cdot \frac{\theta_D}{\theta_A} - \frac{1 - F_D(s_H)}{\phi_D^2}\right]$$

$$s_H^{'}(\phi_D) = \frac{\left(\frac{d\left(\frac{\phi_D}{\phi_A}\right)}{d\phi_D}\right)}{\left(\frac{d\left(\frac{f_A(s_H)}{f_D(s_H)}\right)}{ds_H}\right)} > 0$$

Therefore the objective function is also differentiable with respect to ϕ_D on this open interval and after some manipulation we can express the first-order condition with respect

to ϕ_D (with ϕ_A, s_L, and s_H defined implicitly) as:

$$(E_A[s|s_L < s < s_H] - s_L)\cdot\left(\frac{F_A(s_H) - F_A(s_L)}{\phi_A^2}\cdot\frac{\theta_D}{\theta_A}\right) - (E_D[s|s > s_H] - s_L)\cdot\left(\frac{1 - F_D(s_H)}{\phi_D^2}\right)$$

As $\phi_D \rightarrow^+ \frac{K}{\theta_A\cdot\left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right)+\theta_D}$ it follows that $s_H(\phi_D) \rightarrow^- \overline{s}$ and the right, directional derivative of the objective function with respect to ϕ_D is:

$$(E_A[s|s_L^* < s < \overline{s}] - s_L^*)\cdot\left(\frac{1 - F_A(s_L^*)}{\phi_A^2}\cdot\frac{\theta_D}{\theta_A}\right) > 0$$

Therefore $\phi_D^* = \frac{K}{\theta_A\cdot\left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right)+\theta_D}$ cannot be an optimum, $\phi_D^* > \frac{K}{\theta_A\cdot\left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right)+\theta_D}$ and $s_H^* < \overline{s}$. Finally, we note that when $\phi_D \rightarrow^- \underline{\widehat{\phi}}_D$, it follows that $s_H(\phi_D) - s_L(\phi_D) \rightarrow^+ 0$. Therefore, there exists a critical point to the constrained maximization problem on the interior of the interval $\left[\frac{K}{\theta_A\cdot\left(\frac{f_A(\overline{s})}{f_D(\overline{s})}\right)+\theta_D}, \underline{\widehat{\phi}}_D\right]$. ■

Lemma 5 *The solution function $q_D^*(s) = 1$ on some interval $[s_H^*, \overline{s}]$ where $s_L^* < s_H^* < \overline{s}$.*

Proof. Suppose next that $s_L^* = s_H^* = s^*$. This implies that all student from group D with scores above $s^*$are admitted with probability 1 ($q_D^*(s) = 1\ \forall s \geq s^*$) and that the students from A with scores in this range admitted with probability weakly less than 1 ($q_A^*(s) = \frac{f_D(s)}{f_A(s)}\frac{\phi_A^*}{\phi_D^*} \leq 1\ \forall s \geq s^*$). If there was a unique $\widehat{\phi}_D$ such that $s_{\frac{A}{D}}(\phi_D) = s_D(\phi_D) = s^*$ then as $\phi_D \rightarrow^- \widehat{\phi}_D$ the left, directional derivative of the objective function with respect to ϕ_D is:

$$-(E_D[s|s > s^*] - s^*)\cdot\left(\frac{1 - F_D(s^*)}{\widehat{\phi}_D^2}\right) < 0$$

Therefore, this cannot be an optimum and $s_L^* > s_H^*$. A different argument is required in the absence of uniqueness. Consider any $\widehat{\phi}_D$ such that $s_{\frac{A}{D}}(\widehat{\phi}_D) = s_D(\widehat{\phi}_D) = s^*$. Because both functions are continuous, $\exists\ \epsilon$ such that $\forall \phi_D \in (\widehat{\phi}_D, \widehat{\phi}_D - \epsilon)$ either (1) $s_{\frac{A}{D}}(\phi_D) < s_D(\phi_D)$ or (2) $s_{\frac{A}{D}}(\phi_D) \geq s_D(\phi_D)$. If (1) holds then as as $\phi_D \rightarrow^- \widehat{\phi}_D$ the left, directional derivative exists and is strictly negative by the argument above. If (2) holds then the objective function is equal to $E_D[s|s \geq s_D(\phi_D)]$ which is strictly decreasing in ϕ_D in this interval. It follows that $s_L^* > s_H^*$. ■

Therefore, an optimal no-stereotype policy takes the form:

$$q_D^*(s) = \left\{ \begin{array}{c} 0 \; if \; \underline{s} \leq s < s_L^* \\ \frac{f_A(s)}{f_D(s)} \cdot \frac{\phi_D^*}{\phi_A^*} \; if \; s_L^* \leq s < s_H^* \\ 1 \; if \; s_H^* \leq s < \overline{s} \end{array} \right\}$$

$$q_A^*(s) = \left\{ \begin{array}{c} 0 \; if \; \underline{s} \leq s < s_L^* \\ 1 \; if \; s_L^* \leq s < s_H^* \\ \frac{f_D(s)}{f_A(s)} \cdot \frac{\phi_A^*}{\phi_D^*} \; if \; s_H^* \leq s < \overline{s} \end{array} \right\}$$

This proves Proposition 2. We can further say that $(s_L^*, s_H^*, \phi_A^*, \phi_D^*)$ are a (possibly unique) solution to the first-order condition and constraints for the Kuhn-Tucker program:

$$\frac{1}{\theta_D} \cdot \frac{1 - F_D(s_H)}{(\phi_D)^2}(E_D[s|s \geq s_H] - s_L) = \frac{1}{\theta_A} \cdot \frac{F_A(s_H) - F_A(s_L)}{(\phi_A)^2}(E_A[s|s_L \leq s \leq s_H] - s_L)$$

$$\phi_D \cdot \theta_D + \phi_A \cdot \theta_A = K$$

$$\frac{f_A(s_H)}{f_D(s_H)} \cdot \frac{\phi_D}{\phi_A} = 1$$

$$\frac{1 - F_D(s_H)}{\phi_D} + \frac{F_A(s_H) - F_A(s_L)}{\phi_A} = 1$$