College Diversity Policy and Investment Incentives*

Thomas Gall†, Patrick Legros‡, Andrew F. Newman§

May 2019

Abstract

We study college diversity policies in the presence of local peer effects and pre-college investments. If students are constrained in the side payments they can make within peer networks, the free market allocation displays excessive segregation and investment disparity compared to the first-best. Effective diversity policy must overcome market forces both within and across college boundaries, combining *admission* and *association* policies. When based on achievement, policies can increase aggregate investment and income, reduce inequality, and increase aggregate welfare relative to the market outcome. They may also be more effective than cross-subsidization schemes.

Keywords: Matching, misallocation, nontransferable utility, multidimensional attributes, affirmative action, segregation, education, peer effects, inclusion.

JEL: C78, I28, J78.

---

*Some of the material in this paper was circulated in an earlier paper "Mis-match, Re-match and Investment" which this paper now supersedes. We are grateful for comments from Chris Avery, Roland Benabou, Steve Durlauf, Glenn Loury, John Moore, Andy Postlewaite, and seminar participants at Amsterdam, Boston University, Brown, Budapest, UC Davis, Essex, Frankfurt, Northeastern, the Measuring and Interpreting Inequality Working Group at the University of Chicago, Ottawa, Penn, ThReD 2009, and Yale. Gall thanks DFG for financial support (grant GA-1499). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7-IDEAS-ERC) / ERC Grant Agreement n^0 339950.

†University of Southampton, UK

‡Université Libre de Bruxelles (ECARES), Northeastern University and CEPR

§Boston University and CEPR

1 Introduction

While student diversity in higher education is a goal embraced by many college administrators and policy makers, achieving it has engendered both controversy and challenge. Doubtless a main source of the public controversy is diversity policy's redistributive effect. But there are also concerns about efficiency: favoring certain groups for admission to college may undermine their or other group's incentives to invest at earlier educational stages. Diversity policy may be caught in a classic equity-efficiency tradeoff.

What is more, many have argued that existing policies have been unsuccessful or even counterproductive in achieving the desired level of diversity. Some have expressed frustration that despite generous admission and financial aid polices, the underprivileged have shied away from their (often elite) institutions. Others note that segregation *within* the college gates may be undermining the very ends college diversity is meant to achieve.[1]

By providing a theoretical examination of diversity policy, this paper shows that concerns about efficiency may be overwrought: if students are constrained in the degree to which they can make side payments to their peers, laissez-faire markets need not lead to efficient outcomes either: they will tend toward excessive segregation and distorted investment. Moreover, the dual frustrations are connected: segregation within colleges can induce underprivileged students to stay away. Appropriately designed policies must address the connection between these conundra head on; if they do, they can lead to improvements in diversity, investment, and welfare.

At the heart of the analysis is a role for peer networks in the college experience. They matter both for what students learn while there, and what they earn afterward. They are also a central arena for the formation of long term social relationships, including life partners. These networks are often small — much smaller than the university one attends. If students can freely *associate* within their college, peer networks consisting of single types can feasibly form, even if the college is populated by a full array of types. Thus while the admissions policy of a university may go some way toward accomplishing a given diversity objective, market forces may continue to

[1]Harvard president Drew Faust expressed the typical sentiment: "Simply gathering a diverse mixture of extraordinarily talented people in one place does not in itself ensure the outcome we seek. Everyone at Harvard should feel included, not just represented in this community." (Faust, 2015).

exercise their influence within its boundaries, enabling segregation to persist.

A well-designed diversity policy therefore must address two issues: bring within college walls a diverse population of students and, second, induce diverse peer networks. As we show, policy interventions on both dimensions are jointly necessary to overcome the market forces gravitating toward segregation. Indeed, we will provide a neutrality result (Proposition 2), which states that the imposition of either an admission rule or an association rule in the absence of the other simply replicates the free-market outcome.

Our model considers the following environment. Colleges are arenas for the acquisition of human capital, and, to make our points starkly, this process is driven entirely by local peer effects.[2] At the time they are admitted to college, agents have attributes that reflect their *background* (privileged or underprivileged) and their earlier education *achievement* (high or low). Privilege and high achievement increase both one's own and one's peers' payoffs to attending college. We assume that these local peer effects are strongest when peers have diverse backgrounds. While background is exogenous, achievement is the result of a prior investment. Hence the model is one of creating links among individuals that have multi-dimensional types where some characteristics are endogenous. As far as we know there is no work looking at the role that non transferability plays in such an environment nor what would be the effects of different diversity policies.

Under NTU, the laissez-faire outcome within the college is full segregation in achievement and background within college walls, independently of the admission rule. This implies that the equilibrium choice of investment by individuals is independent of college compositions, or of policy interventions at the admission level.[3] For instance, colleges segregated by characteristics will lead to the same aggregate outcome as colleges admitting students of all characteristics. This also implies that incentives to invest are distorted with respect to a hypothetical "first best" situation, which could be achieved if every agent had unlimited amounts of wealth to make side payments. In

[2]In the Conclusion, we discuss extensions of our analysis to the case in which colleges vary in the inherent quality of their faculty or facilities. Free market outcomes will be little changed; policy analysis will be more subtle.

[3]This modeling strategy frees the analysis from the confounding effects of informational constraints, search frictions or widespread externalities. Indeed, the only frictions in our model are the ones already discussed that inhibit students from making side payments; in particular everyone has full information about each others' types and the payoffs generated from matches, as well as rational expectations about the frequency of attributes (and therefore of different types of matches) in the economy.

general, free-market returns to college for the underprivileged will be low, giving them minimal incentives to invest. The privileged may also have lower incentives to invest than in the first-best situation. But there are also cases in which their incentives are distorted the other way, with very high market returns creating high investment incentives, in which case, the free market situation may be characterized by *over-investment at the top and under-investment at the bottom* (OTUB).

Colleges control admission and can therefore determine the characteristics of students within their gates. But they can also put constraints on how local networks are created among admitted students. We contrast two such association rules. Under *free association*, colleges make no attempt to intervene in students' formation of their local networks. Under *random association*, students cannot choose their local network and are instead linking with other students in proportion to their representation in the college.

Under random association, segregation within the college is precluded if a diverse student body is admitted. But universities could render the random association rule toothless by being selective about the type of students admitted. Hence, association rules by themselves cannot change equilibrium outcome. A combination of constraints on admission – inducing a diverse enough body of students – and association – inducing heterogeneous local networks is necessary for a change in how individuals interact in colleges and generate peer effects.

For policy interventions, we consider two types that impose random association within any college but vary by their constraints on admission (gatekeeping) rules college can use.

We first consider an "affirmative action" policy, which is defined as one that conditions the priority for admission given to an underprivileged on achievement. Affirmative action rewards underprivileged high achievers with access to privileged high achievers, encouraging the underprivileged; at the same time, the privileged are discouraged. The former effect dominates the latter, so that affirmative action generates higher aggregate investment and human capital, and less inequality, than the free market. In fact, aggregate investment under affirmative action tends to exceed that in the first best. Numeric computations indicate that our affirmative action policy can come very close to the optimal re-matching policy.

By way of comparison, and to underscore the economic forces at work, we

then consider "achievement-blind" policies that only focus on replicating the diversity of backgrounds in the population (to be sure, in the U.S. at least, this sort of policy has been largely confined to primary and secondary schools – e.g., busing – rather than higher education). This type of policy guarantees low achievers a "good" match, and high achievers a "bad" one, with sufficient probability as to significantly depress investment incentives. In our model, where achievement contributes significantly payoffs, it is unlikely to have any benefit, and indeed will typically underperform the free market.

Literature

Our model, based on non-transferability in surplus and resulting mismatches in peer groups, leads to novel positive and normative insights, and as such complements other analyses of diversity policy based on imperfections such as search frictions or statistical discrimination.

The literature on college and neighborhood choice (see among others Bénabou, 1993, 1996; Epple and Romano, 1998) typically finds too much segregation in types, often because of widespread externalities (see also Durlauf, 1996*b*; Fernández and Rogerson, 2001), thereby providing a possible rationale for rematch (called "assocational redistribution" in Durlauf, 1996*a*). When attributes are fixed, aggregate surplus may be increased by bribing some individuals to migrate, as in de Bartolome (1990)'s model where there is too little segregation in the free market outcome. Fernández and Galí (1999) compare market allocations of college choice with those generated by tournaments: the latter may dominate in terms of aggregate surplus when capital market frictions lead to non-transferability. They do not consider investments before the match. We complement this literature by focusing on local peer effects as the source of externalities, and by showing that they generate widespread externalities in the form of investment incentives and distribution of individual's human capital.

Rematch has occasionally been supported on efficiency grounds when there is a problem of statistical discrimination (see Lang and Lehman, 2011, for a survey of the theoretical and empirical literature). Coate and Loury (1993) provide a formalization of the argument that equilibria, when underinvestment is supported by "wrong" expectations, may be eliminated by affirmative action policies (an "encouragement effect"), but importantly also point out a possible downside ("stigma effect").

If multiple equlibria are at play, one might expect that, after a rematch policy had been in place for a while, the benefits would persist if it were subsequently removed. This seems inconsistent with empirical observations for colleges: suspending affirmative action policies have often triggered reversion to the pre-policy status quo.[4] Since evolving beliefs are not part of our NTU framework, our model easily explains this observation.

Existing work tends to evaluate the performance of policies with respect to the objective of colleges. E.g., Fryer, Loury and Yuret (2008) evaluate whether a color-blind policy is a better instrument for increasing enrollment of students from a certain background than a color-sighted policy, or the effect of investment of the target group, but do not assess possible general equilibrium effects, e.g., on the group that is not targeted, the privileged, which is a necessary step towards evaluating the effects on inequality or aggregate variables like output or earnings, one of the questions we analyze in this paper.

The theoretical literature on matching (see Chiappori, 2017 or Chade, Eeckhout and Smith, 2017 for more thorough reviews) has illustrated that the composition of groups may be significantly affected by non-transferabilities: while groups may have a diverse composition when a full price system exists, they will be segregated when such a price system is lacking.[5] If the characteristics of matched partners are exogenous, and partners can make non-distortionary side payments to each other (transferable utility or TU); there is symmetric information about characteristics; and there are no widespread externalities, stable matching outcomes maximize social surplus: no other assignment of individuals can raise the economy's aggregate payoff.

Even if characteristics are endogenous, under the above assumptions rematching the market outcome is unlikely to be desirable (Cole et al., 2001; Felli and Roberts, 2016). Peters and Siow (2002) and Booth and Coles (2010) let also agents invest in order to increase their attribute before matching in a marriage market with strict NTU. Booth and Coles (2010) compare different marriage institutions in terms of their impact on matching and investments.

[4]Orfield and Eaton (1996) report an increase in segregation in the South of the U.S. in districts where court-ordered high school desegregation ended, (see also Clotfelter et al., 2006 and Lutz, 2011). Weinstein (2011) finds increased residential segregation as a consequence of the mandated desegregation.

[5]Economists are well aware, at least since Becker (1973), that under NTU the equilibrium matching pattern will differ from the one under TU, and need not maximize aggregate surplus (see also Legros and Newman, 2007 for the case of partially transferable utility.)

Peters and Siow (2002) find that allocations are constrained Pareto optimal (with the production technology they study, aggregate surplus is also maximized), and do not discuss policy. The result of Peters and Siow (2002) has been challenged by Bhaskar and Hopkins (2016) who show that, except in special cases, investments are not first-best when individuals on both sides of the market invest and the surplus is not perfectly transferable. We obtain a similar result in our model, but our focus is on the static (matching) and dynamic (investment) effects affirmative action policies play in environments with non-transferabilities. Gall, Legros and Newman (2006) analyze the impact of timing of investment on allocative efficiency.

Nöldeke and Samuelson (2015) provide a general analysis of matching with non transferability and investment prior to the match. Several other studies consider investments before matching under asymmetric information (see e.g., Bidner, 2014; Hopkins, 2012; Hoppe, Moldovanu and Sela, 2009), *mainly* focusing on wasteful signaling, but not considering rematch policies. Finally, that literature assumes that matching depends only on realized attributes from investment, ignoring therefore the fact that both the initial background as well as the realized attribute may matter for sorting.

2 Model

Consider a market for college populated by a continuum of students with unit measure. Students may differ in their educational *achievement* $a \in \{h, \ell\}$ (for high and low) and their *background* $b \in \{p, u\}$ (for privileged and underprivileged). The set of attributes is

$$\mathcal{A} \equiv \{\ell u, hu, \ell p, hp\}.$$

The distribution of attributes in the economy is q with $\sum_{ab \in A} q(ab) = 1$. Student s has a wealth endowment w_s. In the NTU case $w_s = 0$ (positive, but small values for w_s will not make a difference to our analysis). We will also consider the idealized first best case where w_s is "large" for all agents, as well as the case where only privileged agents have wealth sufficient for making transfers.

Individual background is given exogenously, while achievement is a consequence of a student's investment in education before entering college. Achiev-

ing h with probability e requires an investment in education of e at individual cost $e^2/2$. When entering college, students are fully characterized by their *attributes* ab and their wealth.

Colleges

Students choose to attend one of n colleges. An *admission rule* of a college c is a distribution q_c over A; when $q_c(ab) = 0$, the college does not admit students of attribute ab, but when $q_c(ab) > 0$, the college admits ab students to this proportion. Imposing constraints on college admission rules is one possible instrument of diversity policies.

Once admitted to a college, students interact with their peers, socially and in the accumulation of human capital. These social interactions affect students' payoffs, which are the life time earnings students expect to obtain given their peer group, and the future benefits that these social connections will generate (like referrals for jobs). The pattern of social interactions, i.e. the social network within college is described by the probabilities $p_c(ab, a'b')$ that a student with attribute ab in college c interacts with a student with attribute $a'b'$. These probabilities are endogenous and reflect a choice of *association rule* by the college: for instance, if students can freely choose their roommates, one may get segregation and $p_c(ab, ab) = 1$ for each ab, but if students are assigned randomly into dorms $p_c(ab, a'b') = q_c(ab)q_c(a'b')$. The social network in college given by $p_c(ab, a'b')$ must be consistent with the distribution of characteristics of admitted students.

The network interactions described by $p_c(ab, a'b')$ can reflect both the actual interactions of members of the social network in college or the anticipated probabilities of encountering some significant other in college. In the first interpretation $p_c(ab, a'b')$ denotes the frequency or intensity of interactions; in the second one matching probabilities as in a search and matching or marriage market model. We will refer to $p_c(ab, a'b')$ as the social network.

Definition 1. p_c is *consistent* given q_c if the following two conditions hold.

$$(i)\forall ab,\ q_c(ab) > 0 \Rightarrow \sum_{a'b' \in A} p_c(ab, a'b') = 1,$$

$$(ii)\forall (ab, a'b'),\ p_c(ab, a'b')q_c(ab) = p_c(a'b', ab)q_c(a'b').$$

In general, we expect admission rules q_c to be easier to implement than

association rules p_c. While limiting admission to students with particular attributes (i.e., using $q_c(ab) = 0$ for some attribute ab) is easy to arrange, once a student is admitted, it may be difficult to prevent him or her to interact or not interact with other students. Hence, our setup accounts for the possibility of segregation of social groups within a diverse college, even if admission is subject to affirmative action and leads to a diverse student body. Segregation can be present in social clubs, dormitories, groups of friends, or even in classrooms, if students are free to choose. For instance, Cicalo (2012) reports significant segregation within classrooms attended by law students, where wealthy students sat at the back of the room while poorer students, who often benefited from an affirmative action admission policy, sat at the front (see also Carrell et al., 2013.) Nevertheless, colleges can put in place measures (for roommates' allocation, tutoring or class attendance) that will increase the probability that students of different attributes interact.

Finally, within colleges monetary transfers $t_c : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ may be put in place, where $t_c(ab, a'b')$ is the payment made by a student with attribute ab who interacts with a student with attribute $a'b'$; if negative, $t_c(a'b', ab)$ is the payment received by $a'b'$ when interacting with ab. Such transfers could be centralized by the college; for instance, if the college anticipates social network interactions p_c, the college could collect a payment of $\sum_{a'b'} t_c(ab, a'b')p_c(ab, a'b')$ from students with attribute ab and to give an amount $t_c(a'b', ab)$ to students with attribute $a'b'$ when they interact with a student with ab (e.g., if $a'b'$ tutor ab students).

These transfers are subject to limited liability:

$$\forall a_s b_s,\ \sum_{a'b'} t_c(a_s b_s, a'b')p_c(a_s b_s, a'b') \leq w_s \text{ and } \forall a'b',\ t_c(a'b', a_s b_s) \geq -w_s.$$

A college c can now be defined by the 3-tuple (q_c, p_c, t_c): q_c describes the distribution of admitted students' attributes, while (p_c, t_c) describes the within-college interactions and transfers.

Definition 2. A college (q_c, p_c, t_c) is *admissible* if p_c is consistent given q_c and t_c satisfies limited liability.

Payoffs

All students are assumed to receive zero if they attend no college. The (positive) payoff for each student in an interaction with another student in college when their attributes ab and $a'b'$ is given by:

$$y(ab, a'b') = f(a, a')g(b, b').$$

The output y is the combined market value of human capital $f(a, a')$, taking as inputs individual cognitive skills acquired before the match, and network capital $g(b, b')$, capturing peer effects such as social networks, role models, or access to resources: the marketability of one's human capital depends on the social connections formed at college; or the cost of acquiring human capital at college depends on one's own as well as one's peers' background attributes; or the social environment at college amplifies or depresses the value of individual human capital, or its perception by the market.

Though human capital accumulation depends on one's own characteristics directly as well as through interactions with other students, we will focus on the latter aspect. Letting individual payoffs depend also on the student's attribute, as in the specification $y(ab, a'b') = h(ab) + \hat{f}(a, a')\hat{g}(b, b')$ for some function $h(ab)$, would not alter our main results.

We assume that:

$$f(h, h) = 1,\ f(h, \ell) = f(\ell, h) = 1/2,\ f(\ell, \ell) = \alpha,$$
$$g(p, p) = 1,\ g(p, u) = g(u, p) = \delta,\ g(u, u) = \beta,$$

with

$$\alpha \geq 0,\ \delta < 1, \beta \in [\delta/2, \delta]. \tag{1}$$

As α is non-negative, $f(\cdot, \cdot)$ has increasing differences, consistent with usual complementarity assumptions for production functions. By contrast, the network effects function $g(\cdot, \cdot)$ has strictly decreasing differences on the domain $\{u, p\}$ (that is, $g(u, p) - g(u, u) > g(p, p) - g(p, u)$) if $\delta - \beta > 1 - \delta$, or

$$2\delta > 1 + \beta. \tag{DD}$$

That is, δ captures the desirability of diversity in peer groups: the higher δ is, the more likely that (DD) is satisfied, hence that integration in peer

groups is total surplus enhancing. The parameter β reflects the background gap $g(p,p) - g(u,u)$ between the privileged and underprivileged, the lower β the higher the gap.

We will assume throughout the paper that (DD) holds.[6] There are many reasons for why diversity in backgrounds may indeed be desirable. For instance, when the privileged have preferential access to resources, distribution channels, or information, the benefit of having a peer with a privileged background will be lower for a privileged student, since there may be replication rather than complementarity of information. Furthermore, exposure to peers of a different background enables a student later to cater to customers of different socio-economic characteristics, for instance through language skills and knowledge of cultural norms. Finally, meeting peers of different backgrounds will expose students to methods of problem-solving, equipping them with a broader portfolio of heuristics they can draw on when employed in firms (following the argument by Hong and Page, 2001).[7]

2.1 Timing

The timing in the model economy is as follows.

(1) The diversity policy, if any, is put in place.

(2) Agents choose a non-contractible investment e. Given an investment e, the probability of achievement h is e and of achievement ℓ is $1 - e$.

(3) Achievement is realized and is publicly observed.

(4) Agents match into colleges, who compete in admission and association rules, possibly constrained by the diversity policy.

(5) Within the college agents choose which peer(s) to interact with, governed by the association rule.

(6) Once social interactions are established, payoffs are realized and accrue to the agents.

[6] An earlier version of the paper discussed alternate assumptions on the output function.

[7] Throughout we assume that students perceive the payoff function correctly. It is conceivable that in reality they underestimate the value of diversity; for instance the "true" payoff $\hat{y}(\cdot,\cdot)$ could satisfy $\hat{y}(ap,a'u) > \hat{y}(ap,a'p)$, while for the perceived payoff $y(ap,a'u) < y(ap,a'p)$ as specified above. Another interpretation is that some sort of dynamic inconsistency, like hyperbolic discounting, leads them to behave as if they have the assumed preferences. In either case, the market outcome and positive effects of policy are unchanged, but the case for policy intervention becomes arguably even more compelling.

2.2 Equilibrium

If a student of attribute ab is admitted, that is if $q_c(ab) > 0$, this student has expected payoff

$$u(ab|c) = \sum_{a'b' \in A} p_c(ab, a'b')(y(ab, a'b') + t_c(ab, a'b')).$$

We now define a market equilibrium when the distribution of attributes in the economy is given by q, and consider the overall equilibrium when investment incentives are taken into account below.

In equilibrium, a student with attribute ab chooses a college from the admissible set C^* to maximize her expected utility *conditional* on being admitted. Let $I_c(ab)$ be an indicator, taking value 1 if $q_c(ab) > 0$ and taking value 0 if $q_c(ab) = 0$. When the set of colleges is C^*, a student has indirect utility $v(ab|C^*)$:

$$v(ab|C^*) \equiv \max_{c \in C^*} I_c(ab)u(ab|c), \tag{2}$$

Definition 3. A *college market q-equilibrium* is a set C^* of admissible colleges (q_c, p_c, t_c) together with college enrollments k_c that the following conditions hold:

(i) **Feasibility:** $\forall ab \in A$, $\sum_{c \in C^*} q_c(ab)k_c = q(ab)$.

(ii) **Stability:** there is no admissible college that guarantees all admitted students strictly higher payoffs than their equilibrium payoffs:

$$\forall c \text{ admissible}, \ \{ab|q_c(ab) > 0\} \cap \{ab|u(ab|c) \leq v(ab|C^*)\} \neq \emptyset.$$

Stability reflects competition of colleges for students, by tailoring both admission and association rules to students' demand. In equilibrium, no college can attract a positive measure of students from another college using an admissible combination of admission and association rules and tuition fees. That is, colleges benefit from attracting more students, e.g. to cover fixed cost, or profit maximisation by colleges through tuition fees.

We will provide a constructive proof of existence of a q-equilibrium,.

Investment. Anticipating payoffs $v(ab|C^*)$ for all possible q-equilibria will affect the investments made by agents before they apply to college. Our assumption that attributes in college are determined by stochastic achievement

realizations of a continuum of agents simplifies matters. Indeed, let individuals be indexed by $i \in [0,1]$, with Lebesgue measure on the unit interval. Without loss of generality, assume that all students $i \in [0,\pi)$ have background p and all students in $i \in (\pi,1]$ have background u. If the aggregate investment level of students with background b is e_b, then, by a law of large numbers, the probabilities to the different attributes ℓu, ℓp, hu, and hp are respectively $(1-\pi)(1-e_u)$, $\pi(1-e_p)$, $(1-\pi)e_u$, and πe_p.

Hence college market equilibrium payoffs only depend on aggregates e_u and e_p, all u individuals face the same optimization problem, and all p individuals face the same optimization problem. Therefore agents of the same background b choose the same education investment e_b. We can thus restrict attention to investment strategies $\mathbf{e} = (e_u, e_p)$ that depend on background only. We denote by $q(\mathbf{e})$ the attribute distribution following $\mathbf{e}$.

The pair $\mathbf{e}$ is an equilibrium investment if there is a college market $q(\mathbf{e})$-equilibrium C^* such that for any $b = u, p$

$$e_b = \arg\max_e ev(hb|C^*) + (1-e)v(\ell b|C^*) - e^2/2.$$

Definition 4. An *equilibrium* is an investment $\mathbf{e}$ and a college market $q(\mathbf{e})$-equilibrium, such that $\mathbf{e}$ is an equilibrium investment for the associated C^*.

Generally there is indeterminacy in the (endogenous) sizes of the colleges k_c: if a college (q_c, p_c, t_c) of size k_c is part of an equilibrium, then two colleges (q_c, p_c, t_c) of of size $k_c/2$ could be formed instead and also be part of an equilibrium. Nevertheless, equilibria *outcomes* are essentially unique in the sense that each attribute obtains the same expected payoff in all possible equilibria.

3 Free Market, Non-Transferabilities and Investment Distortions

Before discussing the positive and normative effects of diversity policies, it is useful to contrast the social interactions and the investment levels arising in a free market where agents have little or no wealth, to an idealized one in which agents have no financial constraints and can therefore effectuate arbitrary transfers of utility to their peers.

3.1 Free Market with Non-Transferabilities

In an environment where individual wealth is sufficiently small so that transfers are not possible, a student obtains payoff $y = f(a, a')g(b, b')$ from $(ab, a'b')$ interactions; the Pareto frontier for an attribute pair $(ab, a'b')$ consists of a single point and $t_c(ab, a'b') = 0$. Our assumptions imply that the payoffs to each student in a pair are given by the following matrix. The free market equi-

Attributes	hp	hu	ℓp	ℓu
hp	1	δ	$1/2$	$\delta/2$
hu	δ	β	$\delta/2$	$\beta/2$
ℓp	$1/2$	$\delta/2$	α	$\alpha\delta$
ℓu	$\delta/2$	$\beta/2$	$\alpha\delta$	$\alpha\beta$

Table 1: Individual payoffs from matching into peer group $(ab, a'b')$

librium allocation without side payments has full segregation in attributes: $p_c(ab, ab) = 1$ for all colleges c with $q_c(ab) > 0$. To see this, suppose a college c admits hp students; in this college hp cannot obtain more than 1 in any combination and will segregate; since $\beta > \delta/2$, hu, if admitted, will also segregate since they cannot attract hp; now, because $\delta < 1$, ℓp, if admitted, will also segregate. This precludes positive equilibrium interaction probabilities between students with attributes ab and $a'b'$ (with $ab \neq a'b'$) because this would violate stability. If a college does not admit hp students, the same argument implies that the other types, if admitted, segregate.[8]

While the free market equilibrium interaction probabilities $p_c(ab, a'b')$ are unique ($p_c(ab, ab) = 1$ for any ab) they are consistent with different allocations of students across colleges: the equilibrium remains silent on where the segregation will occur, across or within colleges. One could argue that because it is difficult for students *to completely avoid* students of different attributes when the student body is diverse, that the likely free market outcome is for colleges to be segregated in attributes.

Equilibrium payoffs are uniquely determined, however:

$$v^0(hp) = 1, v^0(\ell p) = \alpha, v^0(hu) = \beta, v^0(\ell u) = \alpha\beta.$$

[8]Note that segregation will be the case if the underprivileged have wealth $w_u < 1-\delta$ and the privileged have wealth $w_p < \beta - \delta/2$. Then still hp students strictly prefer interacting with hp with probability one to any other match, even when obtaining the maximum transfer, and hu students strictly prefer interacting with hu with probability one to any convex combination of hu and ℓp with the maximum transfer.

Therefore an agent of background b chooses e_b to maximize $e_b v^0(hb) + (1-e_b)v^0(\ell b) - \frac{e_b^2}{2}$ implying that $e_b = v^0(hb) - v^0(\ell b)$, and therefore the equilibrium investment levels are:

$$e_p^0 = 1 - \alpha \text{ and } e_u^0 = \beta(1-\alpha). \tag{3}$$

In the free market equilibrium segregation by background is accompanied by differences between individuals of different backgrounds in outcomes such as investments e_b made before the match or expected returns $r_b \equiv e_b v^0(hb) + (1-e_b)v^0(\ell b)$, which can be interpreted as individual education acquisition at college. We use background outcome gaps e_p/e_u and r_p/r_u to quantify investment and payoff inequality.

3.2 First Best: Free-Market with Full Transferability

Utility is fully transferable if individual wealth is sufficiently high, so that the total output from $(ab, a'b')$ interactions

$$z(ab, a'b') = 2y(ab, a'b') = 2f(a, a')g(b, b')$$

can be shared in a 1-1 fashion, that is when the Pareto frontier for $(ab, a'b')$ interactions is obtained by sharing rules in the set

$$\{x : v(ab) = x, v(a'b') = z(ab, a'b') - x\}.$$

The maximum an individual is willing to transfer is equal to $y(ab, a'b')$, corresponding to life time earnings, which for most is a degree of magnitude larger than college tuition fees. Hence, the case of perfect transferability is an ideal rather than a realistic case.

It is well known that under full transferability agents with the same attribute must obtain the same payoff.[9] Because of equal treatment, there is no loss of generality in defining *the* equilibrium payoff of an attribute $v(ab)$. It is also well-known that the equilibrium under fully transferable utility maximizes total surplus given realized attributes.

To derive the surplus maximizing allocation assume that all students attend the same college c, in which attribute shares equal their population

[9]Otherwise, if one agent obtains strictly less than another this violates stability, as the first agent and the partner of the second agent could share the payoff difference.

shares $q(ab)$. The structure of payoffs and the stability conditions lead to the following observations.

(i) $p_c(hp, \ell u) = 0$: $(hp, \ell u)$ interactions cannot occur in a first best allocation. hp interacting with ℓu students lose more compared to their segregation payoff than ℓu students gain: the average surplus for (hp, hp) interactions is 1, $\alpha\beta$ for $(\ell u, \ell u)$, and $\delta/2 < 1/2$ for $(hp, \ell u)$ interactions, less than what segregating hp students obtain.

(ii) $p_c(hp, \ell p) = 0$ and $p_c(hu, \ell u) = 0$: If students of the same background interact in equilibrium, they also have the same achievement. That is, $(hp, \ell p)$ or $(hu, \ell u)$ interactions cannot occur. This follows from increasing differences of $f(a, a')$.

(iii) If students with a given achievement interact, surplus is higher if backgrounds are diverse, because Condition (DD) is equivalent to $2z(ap, au) > z(ap, ap) + z(au, au)$.

(iv) If $\alpha > \delta - \beta$ then $p_c(hu, \ell p) = 0$: $(hu, \ell p)$ matches are not stable, since the sum of segregation payoffs, $1 + \alpha$, is greater than the total surplus in an $(hu, \ell p)$ match, δ.

(v) If $\alpha < 1 - \delta$ then $p_c(hu, \ell p) > 0$ and $p_c(hp, hu) > 0$ only if $p_c(hu, \ell p) < 0$: surplus is higher when $(hu, \ell p)$ interact and hp segregate, than letting (hp, hu) interact and ℓp segregate: in the former case, total surplus is $2\delta + 2$, compared to $4\delta + 2\alpha$ in the latter case. Hence, any equilibrium exhausts all possible $(hu, \ell p)$ interactions and (hu, hp) will interact only if there is an excess supply of hu students.

The policy discussion will be the most relevant when (hu, hp) are the most desirable but do not arise in the free market. At the same time we would like to allow for $(hu, \ell p)$ matches. For these reasons, we will restrict attention to the set of parameters satisfying the following condition:

$$1 - \delta < \alpha < \delta - \beta. \tag{4}$$

Lemma 1. *Under* (4)*, a first best allocation exhausts all possible* (hp, hu) *interactions, then all* $(hu, \ell p)$ *interactions, and then all* $(\ell p, \ell u)$ *matches, while all other remaining attributes do not interact.*

Figure 1 shows the networks that emerge under full transferability according to Lemma 1. Arcs denote interactions between attributes; a plain arc indicates first priority, a dashed arc the second priority, once the supply of first priority attributes is exhausted, and a dotted arc denotes last priority interactions. The first best allocation specifies social interactions, but leaves open the allocation of students across colleges, as long as individuals interact within colleges with the optimal probabilities $p_c(.)$ (for instance, if there are few ℓp, they will all interact with hu students, and ℓu students could segregate in a different college, since $p_c(\ell u, ab) = 0$ for all $ab \neq \ell u$.)

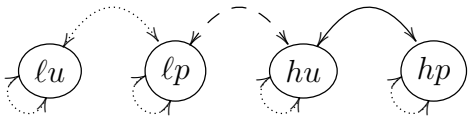

Figure 1: Network interactions with TU

Again the equilibrium definition allows different outcomes in terms of college composition. Some configurations must arise, however: e.g. not all hp students can be segregated, since they should interact with hu students. Hence a top college must admit both hu and hp students. Depending on the distribution q, ℓp students may have to share a college with hu students.

As above investments depend on the market premium for high achievement $v^*(hb) - v^*(\ell b)$. The payoffs for attributes $v^*(ab)$ depend on relative scarcity, which in turn depends on the initial measure of privileged π and achievable surplus $z(ab, a'b')$. For instance, both hp and ℓp students will be less scarce and their payoff lower the higher π. But while hp students' payoff decrease once (when they outnumber hu students), ℓp students payoffs drop twice (when outnumbering first hu, then ℓu students). Hence, the privileged students' return to investment may first fall then increase, which is confirmed in the following statement.

Lemma 2. *Suppose (DD) holds. Under full TU, investment levels e_u^* and e_p^* are non-monotonic in π and vary in opposite directions; e_p^* is U shaped and e_u^* inverted-U shaped.*

If one thinks of the first best outcome as the matching pattern that maximizes total surplus, the following lemma states that the equilibrium of the TU environment indeed leads to a first best allocation. In the proof we show that the payoff difference $v^*(hb) - v^*(\ell b)$ coincides with the social marginal benefit of investment by an individual of background b.

Lemma 3. *The equilibria of the TU environment lead to first best allocations: matching is surplus efficient given the realized attributes, and investment levels maximize ex-ante total surplus net of investment costs.*

3.3 Distortions in Investment

With a price system and unconstrained transfers among students, returns from interactions reflect scarcity: scarce attributes in the market can claim a high share of the total return from network interactions. Therefore the scarcity of the privileged, as measured by π, will affect the returns from college and the incentives to invest in education. By contrast, when there is no possibility of transfer, the returns from social interactions will not reflect scarcity: there will be segregation and therefore the return of an attribute is independent of the attribute distribution, hence of π. This means that privileged students may have lower or higher incentives to invest in the NTU case than in the ideal first-best situation. And indeed, comparing the equilibrium investments e_b^0 under non-transferability to the first-best investment levels e_b^* given in Lemma 2, there is an interval of π for which *privileged agents will over-invest and the underprivileged under-invest* with respect to the first-best. This "over-investment at the top, under-investment at the bottom" (OTUB) outcome starkly illustrates the possible investment distortions that can be brought about by non-transferabilities.

The following proposition offers a more precise characterization of the investment outcomes (comparing the laissez-faire investment levels to the first best ones in the proof of Lemma 2) and illustrated in Figure 2.[10]

Proposition 1. *There are* $0 < \hat{\pi}_0 < \hat{\pi}_1 < \hat{\pi}_2 < 1$ *such that*

- *The underprivileged never over-invest; they under-invest if* $\pi > \hat{\pi}_0$.
- *The privileged under-invest if* $\pi < \hat{\pi}_1$ *and over-invest if* $\hat{\pi}_1 < \pi < \hat{\pi}_2$.

There is both over-investment at the top and under-investment at the bottom of the background distribution, if $\hat{\pi}_1 < \pi < \hat{\pi}_2$.

This result formalizes the idea that imperfect transferability within peer networks can generate not only excessive segregation, a static inefficiency,

[10] In this figure as well as others in the paper we use the parametrization $\delta = 0.9$, $\beta = 0.6$, $\alpha = 0.2$.

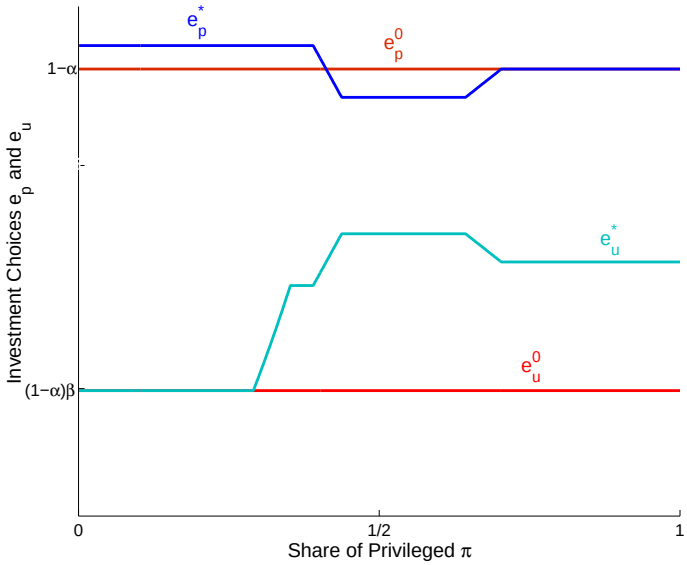

Figure 2: Education investments: NTU (e^0) vs TU (e^*)

but also investment distortions, a dynamic inefficiency. Specifically, the underprivileged will tend to under-invest; as for the privileged, their investment will be insufficient or excessive depending on whether they are a small enough minority. This suggests that the possible discouragement effects on the privileged that diversity policies introduce may sometimes actually be desirable.

In the absence of transfers, payoffs do not have full freedom to adjust. Their role as price-like signals scarcity that can attract investment where it is socially most useful is thereby diminished. When the privileged are scarce, they receive somewhat lower payoffs in the segregated free market outcome than they would in a TU world, where they would be receiving large side payments. Their free market investments are correspondingly lower than they would be under TU.

Excessive segregation also has implications for inequality and polarization, but not necessarily in the "obvious" way. Indeed, computing background gaps as a measure of inequality yields the following corollary.

Corollary 1. *For intermediate and high π, inequality in investments e and in payoffs y is higher under NTU than in the first best.*

Hence, if backgrounds are distributed relatively equally, excessive segregation is accompanied by excessive income inequality. In other instances however, income inequality may be greater in the first best benchmark as scarce attributes are paid their full market price (for instance when π is close

to 0, hp agents obtain $2\delta - \beta$ in the first best, but only 1 under free market).

4 The Positive and Normative Effects of Diversity Policies

Real world policies aim at replicating population measures of backgrounds in colleges, but vary in whether they are applied at the admission stage or within college and in the degree to which they allow colleges to condition on achievement. Specifically, policy can determine the admission and association rules colleges can employ. One could target admission, i.e. the shares of each attribute admitted to a college c, $q_c(ab)$, for instance requiring all colleges to equate them to the population shares. Policy could also target association within college, i.e. the interaction probabilities $p_c(ab, a'b')$. While constraints on admission rules are relatively easy to impose (except for possible legal problems), the matching and mingling of student on campus is much harder to steer. Colleges have a degree of control, however, be it in form of dorm room lotteries (Sacerdote, 2001), or teaching techniques (Cicalo, 2012).

We first show that any diversity policy that only targets admission or only association will be ineffective (the proof is in the appendix).

Proposition 2. *Under NTU, imposing on all colleges exclusively either an admission rule or an association rule will yield segregation as the unique equilibrium outcome for almost all parameter values.*

The reasoning behind Proposition 2 is that any diversity policy at the college gates can be undone by free association within the college, while any association policy within the college walls is rendered toothless, if admission rules allow students to sort across colleges. The latter seems an accurate description of actual outcomes in the U.S. and the U.K. For the former, Cicalo (2012) provides an illustration of the force towards segregation within colleges and describes how students of different backgrounds segregated physically in the classroom under a quota policy in admission. Another example is an experiment sorting freshmen at the United States Air Force Academy into groups with diverse ability, which was undone by the freshmen forming segregated friendship networks (Carrell et al., 2013).

Therefore, to be effective, policies must constrain both the way admissions are made *and* the way students match within colleges.

4.1 Random Association within College

From now on we focus on diversity policies that affect both admission and within college interactions, and that lead to equilibria that differ from the free-market equilibrium. We model policies affecting association on campus in the simplest conceivable way, by letting the diversity policy impose random matching within college. In addition to analytical convenience, this modeling choice also accommodates the current reluctance of most universities to abandon the principle of free association.[11]

Definition 5. Random association implements random matching within the college walls, that is $p_c(ab, a'b') = q_c(a'b')$.

The adoption of such association rules could be in response to social pressure for colleges to show that they are not only admitting a diverse student body, but also that minority or disadvantaged students are truly integrated in the university. Such pressure may lead to the disappearance of fraternities or sororities, the move from self-selection for roommates and assignment to dorms to a lottery system at the university level, or lottery system for the assignment of students to popular courses.

A no arbitrage argument implies the useful result that in a college market equilibrium under random association generically colleges are symmetric with respect to their attribute composition, since otherwise students would find it profitable to migrate to colleges that have higher share of more desirable attributes (details are in the appendix).

Lemma 4 (Symmetric Equilibria)**.** *Under random association, for almost all parameter values, if in an equilibrium the supports of types for two colleges coincide, they have the same distribution of types.*

For college admission we focus on two extreme policies. First, we will consider an "affirmative action" (A) policy, which gives priority to the underprivileged over privileged students in admission, but conditions this priority on achievement. The effect of this policy is to generate mixed-background

[11] Exceptions to this reticence may prevail in the U.S. military academies; going beyond the university context, private firms routinely restrict their employees' ability to free associate, whether by work rules, assignment to teams, restricted meal hours, etc. An earlier version of the paper considered the case that colleges have full firm-like control over social interactions on campus; the results are qualitatively the same, but quantitative effects are stronger, as random matching is not in general the optimal association rule.

networks, while maintaining a fair degree of segregation in achievement, thereby providing a better approximation to the TU benchmark than does the free market.

To underscore the importance of designing policy that takes account of incentives, we contrast the A policy with one that is achievement-blind – a caricature perhaps of actual university diversity policies – ignoring achievement and instead basing admission only on backgrounds, to replicate the population frequencies: the probability that a u interacts with a p is just π.

Because a large part of the social surplus is linked to the achievement element of the attributes, an achievement-blind (B) policy tends to perform worse than either the free market or affirmative action. Studying these polar cases allows some inference on intermediate ones, like scoring policies where a score reflecting both achievement and background determines priority.

4.2 Affirmative Action Policy

We begin with the case where precedence is given for an underprivileged candidate over a privileged competitor if both have the *same* achievement level.[12] Policies of this type are widely used (for instance, reserving places for highly qualified minority students at some *grandes écoles* in France, like Sciences Po Paris, the "positive equality bill" in the U.K. and *Gleichstellung* in the German public sectors). In higher education particular attention seems to be on broadening access to the most selective universities (such as Oxford and Cambridge in the UK and the Ivy League in the U.S.). Therefore we also consider a policy of affirmative action at the top, only awarding priority for the underprivileged high achievers over their privileged counterparts.

Definition 6. Under an *affirmative action policy* (denoted A policy) any underprivileged student with achievement a is guaranteed a place at any college that also admits privileged students with the same achievement a. Under *affirmative action at the top* ($\overline{A}$ policy) any underprivileged student with achievement h is guaranteed a place at any college that also admits privileged students with the same achievement h. Within colleges an association policy is in place that ensures random matching.

[12]This policy, using background only as a tie-breaker if achievement is high, is probably closest to the free market in requiring only a minor intervention. Many different affirmative action policies are also conceivable and could be analyzed in this framework.

That is, an A policy ensures that if $q_c(ap) > 0$ for a college c then also $q_c(au) > 0$. Under both policies no arbitrage has to ensure that students with attribute au strictly prefer their equilibrium college to any other college with $q(au) > 0$ or $q(ap) > 0$, where $a \in \{\ell, h\}$ or just $a = h$. Hence, full segregation can no longer be stable as hu students prefer to be in colleges that have both hu and hp students. The following lemma provides a key property of the college market equilibrium.

Lemma 5. *Under both A and $\overline{A}$ policies all hu and hp will enter colleges with $q_c(hp) = \pi e_p/(\pi e_p + (1-\pi)e_u)$ and $q_c(hu) = 1 - q(hp)$. Under an A policy all ℓu and ℓp will enter colleges with $q_c(\ell p) = \pi(1-e_p)/(\pi(1-e_p)+(1-\pi)(1-e_u))$ and $q_c(\ell u) = 1 - q(\ell p)$. Under an $\overline{A}$ policy all ℓu and ℓp students will segregate across colleges, i.e., $p_c(\ell u, \ell u) = 1$ and $p_c(\ell p, \ell p) = 1$.*

Proof. By Lemma 4 college market equilibria are symmetric: all colleges with the same support of attributes have also the same student composition.

Under full NTU all students prefer to be matched with hp, then hu, then ℓp and then ℓu students. Both hp and hu students will have an incentive to switch to colleges with higher $q_c(hp)$. Hence, no arbitrage implies that $q_c(hp)$ must be constant across colleges. Since ℓb students have no priority over hb students $q_c(hp) > 0$ implies $q_c(\ell b) = 0$ for $b = u, p$. If ℓp and ℓu students have no priority they will segregate as in the free market equilibrium. Otherwise both ℓp and ℓu students will have an incentive to switch to colleges with higher $q_c(\ell p)$ and no arbitrage implies that $q_c(\ell p)$ must be constant across colleges with $q_c(\ell p) > 0$. □

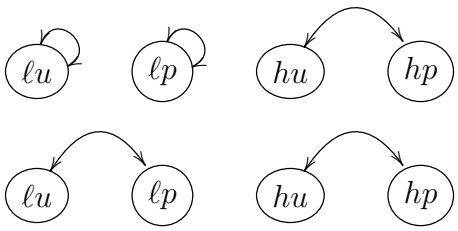

Figure 3: Network interactions under $\overline{A}$ (top) and A (bottom) policies.

Figure 3 shows the corresponding equilibrium networks. The one resulting from an $\overline{A}$ policy is consistent with three types of colleges emerging in equilibrium: colleges who admit only ℓu students, those admitting only ℓp students and finally those admitting hp, hu students; in the last type of colleges random association ensures interactions between hu and hp.

Individual investment depends on the network a student expects to obtain, and thus on relative scarcities. Since privileged high achievers end up interacting with other high achievers only, but of both backgrounds, their expected payoff will be lower than in the free market outcome. Underprivileged high achievers will have higher payoffs than in the free market outcome, because they interact with privileged high achievers. Low achievers' payoffs coincide with the free market under an $\overline{A}$ policy. Therefore the privileged (underprivileged) will invest less (more) than in the free market.

Moving to a broad-based A policy that also integrates ℓ colleges will redistribute payoff from privileged to underprivileged low achievers. Thus the privileged will increase and the underprivileged decrease investments compared to an $\overline{A}$ policy. High achievers of both backgrounds will have higher payoffs, however, because the changed investment improve the pool of high achievers, increasing the proportion of privileged backgrounds. The overall effect is to increase aggregate welfare and output.

The following proposition states this and other properties of aggregate outcomes under affirmative action policies; details are in the appendix.

Proposition 3. *Under both A and $\overline{A}$ policies, compared to the free market*

- *the underprivileged invest more ($e_u^A > e_u^0 > e_u^B$), and the privileged less ($e_p^0 > e_p^A > e_p^B$),*
- *inequality of investments between backgrounds is smaller,*
- *aggregate investment is higher; aggregate output and welfare are higher if diversity is desirable enough.*

Welfare and output, and investment inequality are higher under an A policy than under an $\overline{A}$ policy.

That is, both affirmative action policies only moderately reduce privileged investment and underprivileged investment is boosted compared to the free market, as illustrated in Figure 4 for the A policy (the picture for an $\overline{A}$ policy looks very similar). This is because under an A policy an underprivileged student's expected return from investment is given by the difference of being admitted to a (hu, hp) college rather than to a $(\ell u, \ell p)$ college. Therefore the expected returns to investment are now conditional on integrating in backgrounds if successful. This encourages the underprivileged and discourages the privileged, and, if diversity is desirable (condition (DD) holds),

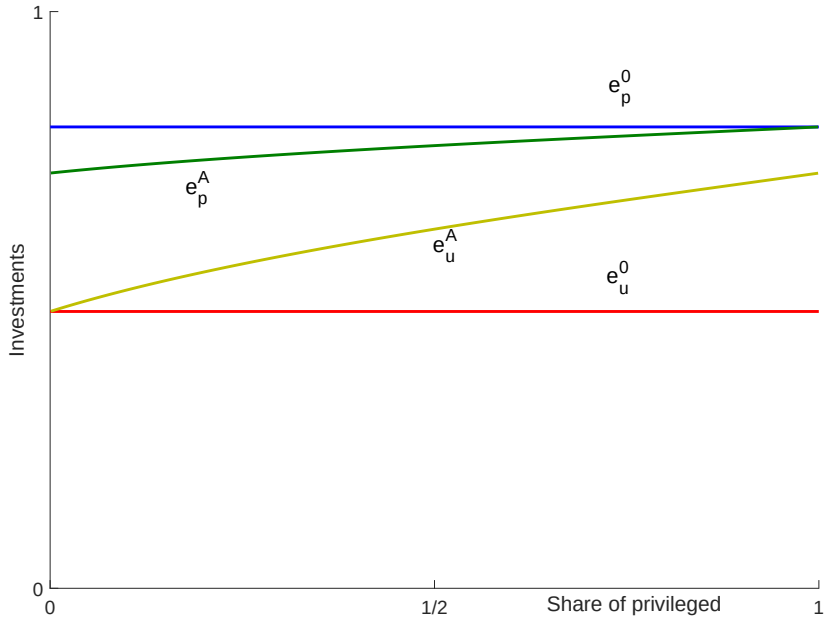

Figure 4: Education investments using an A policy.

the aggregate effect on investment is positive. If diversity is very desirable or backgrounds are distributed unevenly also aggregate output is higher. Perhaps surprisingly, the redistributive effect of an $\overline{A}$ policy on incentives can be substantial: the education gap between backgrounds may reverse if $\delta > 1 - \alpha(1 - \beta)$ and the measure of privileged is close to 1.

This comparative statics exercise assumes that when π varies, both δ, β stay constant, which may be a strong assumption in general.

4.3 Achievement-Blind Admission Policy

The policy we consider here replicates the population distribution of backgrounds in each college, unconditional on achievements.[13]

Definition 7. Under an *achievement Blind policy* (denoted B policy) a college cannot discriminate on the basis of achievement for admission and has

[13]Few real-world policies, particularly in universities, are truly achievement blind; the most prominent examples applied to other arenas, such as "busing" to achieve integration in U.S. primary and secondary schools, or the Employment Equality Act in South Africa, under which some industries such as construction and financial services used employment or representation quotas. Even the post-1968 European university practice of not conditioning admission on achievement beyond the basic requirement of finishing high school, which might appear superficially to be equivalent to an assignment rule that randomly integrates peer groups in background, likely did not operate this way, because the universities were generally permissive of free association within their boundaries.

to admit u students if it admits p students. Within colleges there is random association.

This policy precludes, in equilibrium, the formation of colleges admitting only h students. h students would value such colleges to avoid interactions with ℓu or ℓp students under random association. However, such a segregated college cannot refuse to admit ℓ students, who benefit from interacting with h students. Lemma 4 then implies that colleges offer compositions $q_c(ab)$ equal to population shares in equilibrium. A B policy is thus best understood as a quota policy that departs from the free market outcome of full segregation and randomly reassigns students to match the share of privileged at each college to their population share π.

The following statement describes the resulting assignment of students.

Lemma 6. *Under a B policy all colleges have the same share of both backgrounds, and the same share as the population of students: $q_c(\ell p)+q_c(hp) = \pi$ and $q_c(\ell u) + q_c(hu) = 1 - \pi$. Ex-post probabilities of linking with peers are given by the population shares of attributes: $p_c(ab, a'b') = q(a'b')$.*

The statement is straightforward, but uses a law of large numbers for achievements. Figure 5 describes the possible matches.

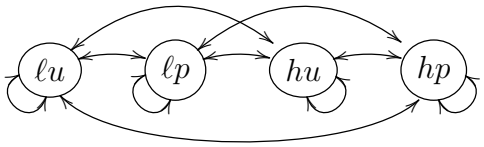

Figure 5: Network interactions under a B policy.

Because this policy allows both (hu, hp) and $(\ell p, hu)$ interactions, it may be beneficial for increasing surplus if, unlike in our model, investment in achievement is not important, for instance when the distribution of types is exogenous. However, in the broader set of cases, investment incentives are likely to be depressed compared to the free market.

The following statement uses Lemma 6 to verify this intuition; details are in the appendix:

Proposition 4. *Investments under a B policy are lower than in the free market outcome for both backgrounds, as are aggregate investment and payoffs. This policy induces both lower payoffs and lower investment inequality between backgrounds measured by the ratio than the free market.*

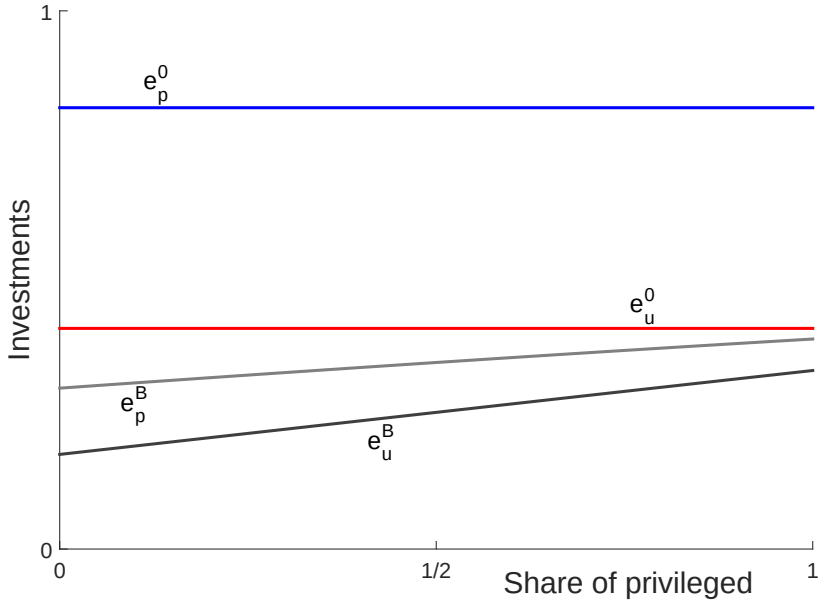

Figure 6: Education investments: B policy (e^B) and laissez-faire (e^0).

That is, a B policy reduces outcome inequality in the economy at the cost of undesirable incentive effects depressing levels of investment and output, see Figure 6.

4.4 Aggregate Effects

The diversity policies considered above differ substantially in terms of how they trade off static and dynamic concerns, that is efficient sorting ex post (when attributes have been realized) and efficient investment incentives by rewarding investments adequately through the match. Policies that emphasize replicating population frequencies of backgrounds (B policies) may do well in terms of the first but will in general fail in terms of the second. Policies that implement admission of students with similar achievement levels forgo some benefits of improving the sorting ex post, since for instance matches $(\ell p, hu)$ will not be realized, but induce high investment incentives, mainly by providing access to mixed colleges for the underprivileged. Figure 7 illustrates the differences in aggregate performance.

Both types of policy tend to decrease inequality in the economy compared to a free market: they decrease the privileged's investment incentives substantially, while the underprivileged's incentives increase with access to better matches. Here investment inequality is also an indicator of social mobility, in terms of the predictive power of parental background on own

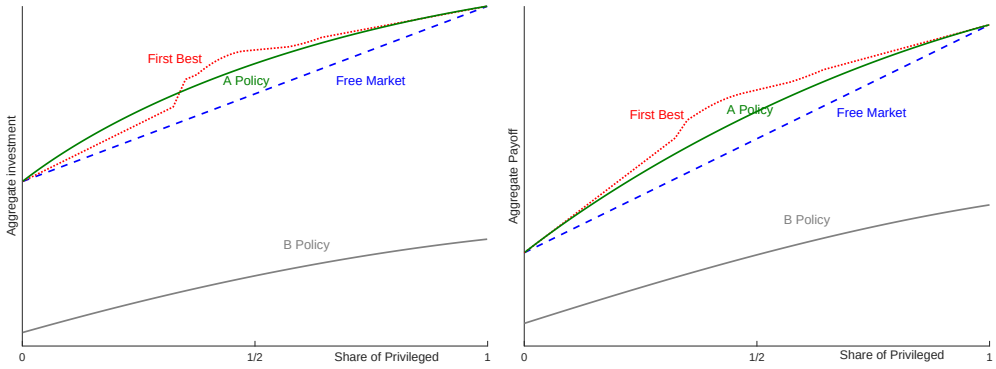

Figure 7: Aggregate investments (left) and aggregate payoff (right).

achievement and payoffs. Figure 8 shows the investment and payoff ratios of privileged to underprivileged.

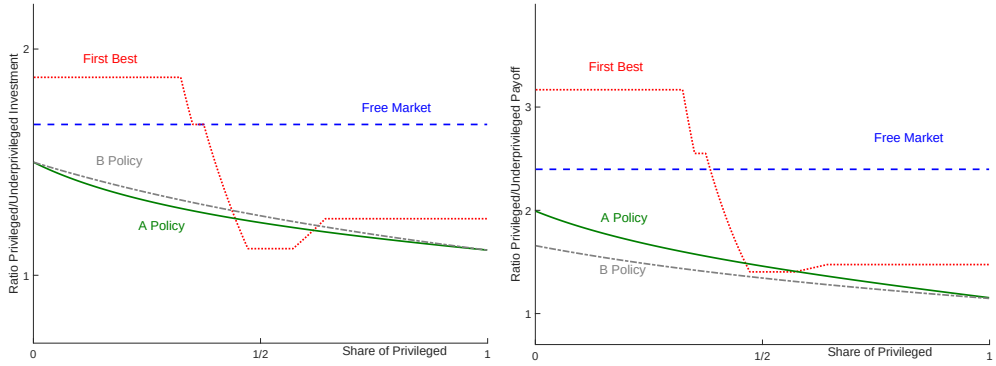

Figure 8: Inequality of investments (left) and payoffs (right).

Our results suggest that policies that ignore achievement, focusing only on background, are likely to be far less effective in improving various aggregate outcome measures, and some of them will do more harm than good. Properly designed achievement based policies, for instance in the form of scoring rules that assign high weight to high attainments, are preferable to those that simply mix in terms of backgrounds, and can be quite effective in improving both aggregate efficiency and equity.

The same conclusions apply if we focus not on outcomes such as output, inequality and investment, but on welfare, measured in aggregate surplus, that is, expected payoff net of investment cost. See Figure 9.

In this figure the A policy clearly dominates the free market under NTU and the B policy. The dominance of A over B in terms of welfare is a general property, but that of A with respect to NTU requires that δ be large enough (as in the figure where $\delta = 0.9$).

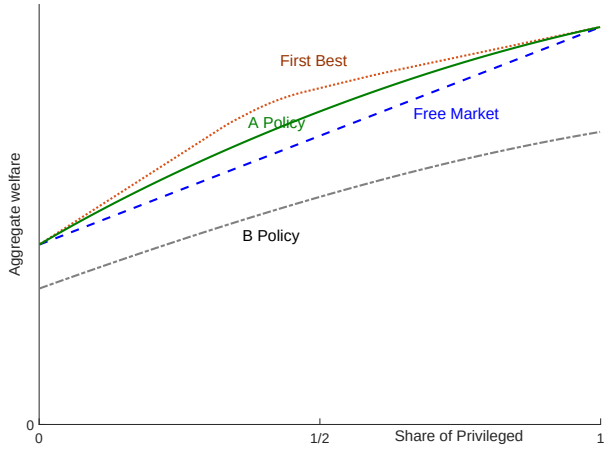

Figure 9: Total Surplus

Proposition 5 (Welfare)**.** *(i) The free market dominates a B policy in terms of total surplus.*

(ii) For each $\pi \in (0,1)$, there is $\hat{\delta}(\pi) < 1$ such that an $\overline{A}$ policy induces strictly higher total surplus than the free market with NTU if $\delta > \hat{\delta}(\pi)$.

(iii) An A policy dominates an $\overline{A}$ policy in terms of total surplus.

4.5 Direct Association Policy

Figure 9 shows that the A policy, relying on random associations, achieves gains in surplus compared to the laissez faire outcome. A social planner who has full control over agents' associations (such as military academies or firms) may achieve even higher surplus. The optimization problem of a planner choosing interaction probabilities $p(ab, a'b')$ subject to feasibility is:

$$\max_{p} \sum_{ab,a'b'} p(ab,a'b')z(ab,a'b') - \pi\frac{e_p^2}{2} - (1-\pi)\frac{e_u^2}{2}$$

subject to incentive constraints: for $b = p, u$:

$$e_b = \sum_{a'b'} \frac{p(hb,a'b')}{\pi_b e_b} y(hb,a'b') - \sum_{a'b'} \frac{p(\ell b,a'b')}{\pi_b(1-e_b)} y(\ell b,a'b'),$$

and feasibility: for $b = p, u$:

$$\sum_{a'b'} p(hb, a'b') + p(hb, hb) = \pi_b e_b \text{ and}$$
$$\sum_{a'b'} p(\ell b, a'b') + p(\ell b, \ell b) = \pi_b (1 - e_b).$$

That is, the set of policies contains all feasible interaction patterns between different attributes, which in turn determine investments, and the optimal solution will achieve a second best in this sense. Recall that an A policy will set $p(hu, hp)$ equal to the population shares and $p(\ell p, \ell p) = 1$ as well as $p(\ell u, hu) = 1$.

The problem above has six control variables and a discontinuous objective function, making the problem hard to solve analytically. Numerical solutions indicate that the direct association policy closely resembles an policy that integrates backgrounds conditional on achievements, as an A policy, but uses directed assignment instead of random assignment within the college, thus exhausting *all* (hp, hu) and $(\ell p, \ell u)$ matches. Simulations indicate that the ability to target interactions significantly improves the welfare properties of the policy: the second best policy, using directed matching, achieves between 90.6% and 97.1% of the gap between first best and free market surplus when $\pi \geq 1/2$, and between 72.6% and 79.6% when $\pi < 1/2$, whereas the A policy only achieves between 44.2% and 81.3% when $\pi \geq 1/2$ and between 40.0% and 66.7% when $\pi < 1/2$ (for $\delta = .9$, $\beta = .6$, and $\alpha = .2$, used for all figures).

5 Partial Transferability

Another remedy to excessive segregation implied by NTU could consist in "bribing" ex-ante some students to re-match. Indeed, while a complete lack of side payments appears to describe well the assignment of pupils to public colleges, at all levels of education there are private colleges that charge tuition fees that may reflect students' academic achievements, for instance by offering scholarships. This introduces a price system for attributes, potentially affecting both the matching outcome and investment incentives. Often such a price system suffers from imperfections, for instance because individuals differ in the financial means at their disposal that can be used to pay tuition fees and some of them face borrowing constraints. As we already

pointed out, since benefits from college are related to lifetime earnings, it is likely that the financial constraint binds for most students.

We introduce the possibility of transfers among students by assuming that agents differ in their wealth levels w_b, depending on their background b. Plausibly, privileged background is associated with higher wealth. As mentioned in footnote 8, for $w_u < \alpha(1-\delta)$ and $w_p < \beta - \delta/2$ our previous analysis goes through unchanged, because hu students cannot compensate hp students enough to depart from the segregated outcome; neither can ℓp's compensate hu's, nor can ℓu's attract ℓp's. Suppose for simplicity that

$$w_p > \delta/2, \text{ and } w_u = 0. \tag{5}$$

This implies that the privileged can compensate the underprivileged, but not vice versa; see Figure 10 for the possible payoffs for some attribute combinations.

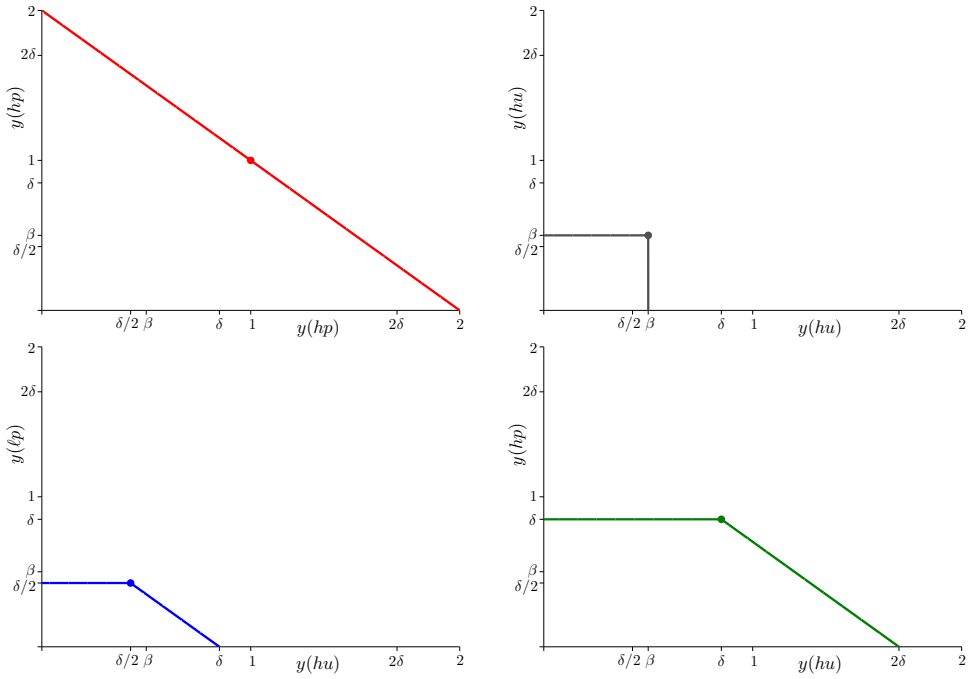

Figure 10: Possible distribution of payoffs in (hp, hp) and (hu, hu) interactions (top) and $(\ell p, hu)$ and (hp, hu) interactions (bottom) when individuals can make lump-sum transfers but the underprivileged face borrowing constraints.

The next statement follows directly from this observation.

Lemma 7. *Under the wealth distribution assumption* (5)*, in equilibrium,*

there are three types of colleges: those composed of ℓu students, those composed of hp students and those composed of $(hu, \ell p)$ students.

Figure 11 shows the resulting college market equilibrium. The underprivileged interact with the privileged, but only in $(hu, \ell p)$, not in (hu, hp) colleges, and elite (hp, hp) colleges are solely populated by the privileged, which seems to resonate well with the evidence.[14] .

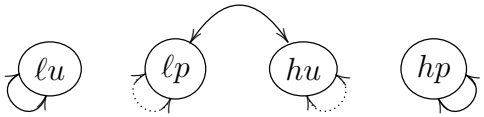

Figure 11: Network interactions in an equilibrium with transfers

Affirmative action policies (combined with random association within the college) will have an effect on the equilibrium networks. If, in line with observations, only the most selective colleges employ affirmative action, with (5) an $\overline{A}$ policy yields (hp, hu) interactions, since hu students have priority in college admission. However, ℓp students would still pay to interact with hu students and thus hu can choose between better peers and better money. If hu are scarce and interact with hp with near certainty, ℓp students prefer segregating to compensating. That is, for high shares of the privileged the $\overline{A}$ policy with partial transferability coincides with the one under NTU.

Lemma 8. *Under the wealth distribution assumption* (5)*, any college market equilibrium under an $\overline{A}$ policy yields colleges with both hu and hp students, and, if the share of privileged π is low enough, colleges with both hu and ℓp students. ℓu students attend segregated colleges.*

As in the case without side payments, an $\overline{A}$ policy encourages investment by the underprivileged, since underprivileged high achievers are rewarded with access to privileged high achievers. By contrast, when side payments

[14]For instance, Dillon and Smith (2013) find evidence for substantial mismatch in the U.S. higher education system, in that students' abilities do not match that of their peers at a college. This mismatch is driven by students' choices, not by college admission strategies, and financial constraints play the expected role: wealthier students, and good students with close access to a good public college are less likely to match below their own ability. Hoxby and Avery (2013) report that low-income high achievers tend to apply to colleges where the average achievement is lower than their own achievement and seem less costly, in contrast to the behavior of high income high achievers (Table 3). They also find that prices at very selective institutions were not higher for the underprivileged than at non-selective institutions, although this does not account for opportunity cost of, e.g., moving. At a theoretical level the outcome is evocative of Epple and Romano (1998).

are possible an $\overline{A}$ policy may encourage investments by students of *both* backgrounds. This is because limited wealth limits competition among ℓp's, giving rents to privileged low achievers. An $\overline{A}$ policy depresses the rent for privileged low achievers, forcing them to compete with privileged high achievers for scarce underprivileged high achievers (when π is intermediate). This effect outweighs the decrease of the privileged high achievers' payoffs, now forced to interact with the underprivileged, so that investment incentives for the privileged increase. For intermediate π this encouragement effect is so strong that the expected payoff ex post of a privileged student is higher under an $\overline{A}$ policy, if diversity is desirable enough (δ sufficiently large).

Proposition 6. *Relative to the free market, an $\overline{A}$ policy induces*

- *higher investment and payoffs for the underprivileged, and lower investment gaps between backgrounds,*
- *higher investment for each background, and for intermediate π also higher payoffs for both backgrounds, if δ is high enough.*

Figure 12 illustrates the change in aggregate outcomes as a function of the proportion of privileged students when colleges use tuition fees.

Until now, we have considered the possibility of transfers between students who are in the same social network, and have shown that an affirmative action policy still has a role to play in generating (hu, hp) interactions, and improving on aggregate variables like output, investment and welfare.

However, because hu students have the right but are not compelled to interact with hp students under affirmative action, and because the privileged have wealth with which to make side payments (perhaps intermediated through universities), there may be incentives for hp's to encourage the hu's to match elsewhere, as well as for ℓp's to attract the hu's. This requires some transfers across networks (from hp's to hu's, who would join (hu, hu) or $(\ell p, hu)$ groups instead of (hu, hp) ones), and the consideration of deviations by coalitions of more than two individuals.[15]

For instance, hp and ℓp students in (hp, hu) and $(\ell p, \ell u)$ colleges could jointly offer side payments to hu students to achieve a rematch into colleges

[15]In practice, such transfers could be effectuated through donations by the hp's (or their families) to the scholarship funds of other peer groups' colleges, as exemplified by the Koch brothers' donations to the United Negro College fund (http://www.thewire.com/politics/2014/07/major-union-blacklists-united-negro-college-fund-for-koch-brothers-relationship/374264/).

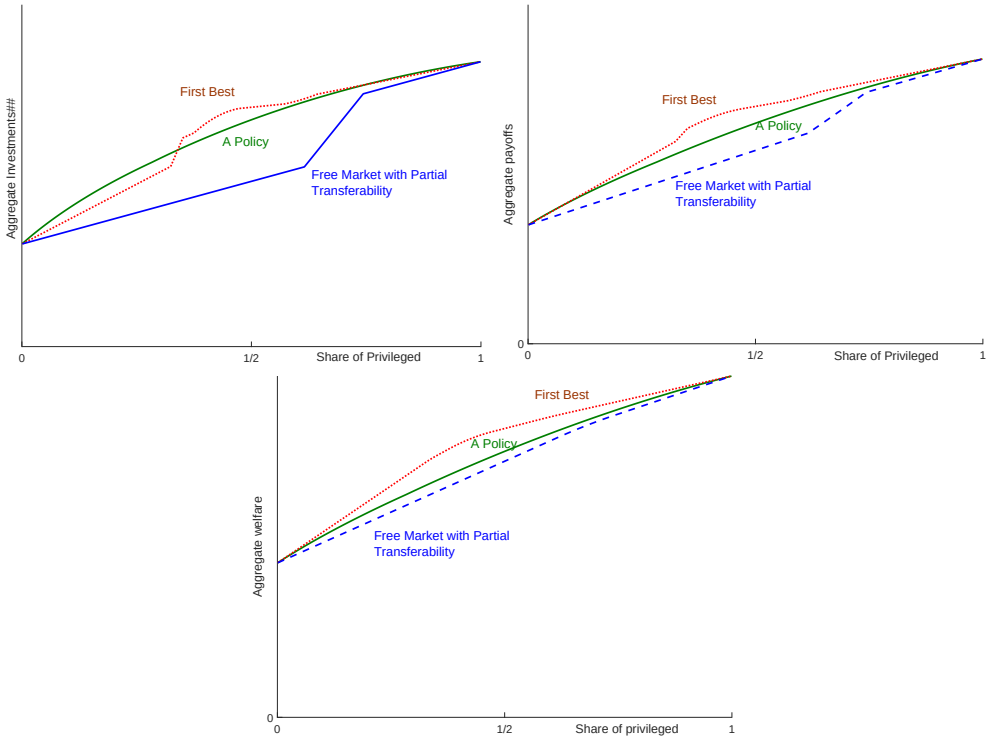

Figure 12: Aggregate investments (left), income (right), and surplus (bottom) when $w_p > \delta/2 - \alpha, w_u = 0$

$\{(hp, hp), (hu, \ell p), (hu, \ell p), (\ell u, \ell u)\}$. Since ℓu students have no priority in mixed groups nor over h students, they do not have to be bought off. hu students would prefer this arrangement if the side payment exceeds $\delta/2$. An hp student would be prepared to pay at most $1 - \delta$ to interact with an hp, and ℓp students would pay at most $\delta/2 - \alpha\delta$ to interact with ℓu instead of hu students. That is, given an $\overline{A}$ policy, an outcome that exhausts all (hp, hu) and $(\ell p, \ell u)$ matches *will not* be stable when

$$\delta < \frac{1}{1+\alpha}.$$

Under this condition, an $\overline{A}$ policy will not lead to (hu, hp) matches but will in fact replicate the free market equilibrium of Figure 11.

But despite the fact that the policy does not seem to have had an effect on *college composition*, it still benefits the underprivileged, increasing their incomes, investment incentives and welfare (in fact in our example, the investment incentives of the u's are higher than they would be if the $\overline{A}$ policy only led to rematch, while the p's have the same investment incentives whether or not the rematch is effected – thus the $\overline{A}$ policy generates higher

aggregate investment than the market outcome whether or not it can be destabilized). Affirmative action may lead to a redistribution of wealth, even if it does not lead to a redistribution of students.

Another category of diversity policies is the use of scholarships, especially for hu's, financed by private endowments or government funds. These try to generate (hu, hp) interactions by giving the hu's sufficient wealth to make the side payment needed to enter a (hu, hp) college. Usually this is a voucher or scholarship, since the wealth given to the hu cannot be spent arbitrarily. Observe however, that if the hp with whom the hu is supposed to be paired does not also receive the side payment (perhaps in the form of his own tuition discount), he will not be willing to match with the hu and will instead segregate with another hp. As in the free market outcome, the result is a preponderance of (hp, hp) matches, along with $(\ell p, hu)$. The outcome is the result of market forces among fully informed rational actors, with only borrowing constraints at play. Scholarships need not be substitutes for transferability within networks.

6 Conclusion

A perceived excess of segregation in the collegiate marketplace has inspired many policy responses as well as much controversy. Starting with a model in which the benefits of higher education accrue through peer networks formed within a college, and in which students have limited means with which to make transfers, we show that the free market will indeed generate excessive segregation. As a consequence, pre-college investments are distorted, with under-investment by the underprivileged and often over-investment by the privileged. These outcomes occur even if students know that they benefit from diversity and even if the benefit is at the level of their (small) peer network: local non-transferabilities are the root of all distortions.

Policy that effectively countervails these market forces must apply both across colleges (admission policy) and within them (association policy). Applying only one or the other can be fully neutralized by student arbitrage in network formation or college choice.

Some Ivy League universities have expressed consternation at their seeming inability to attract as many underprivileged high achievers as they would like, despite offering generous scholarships to the under-privileged (Hoxby

and Avery, 2013). In our model, without transfers to the privileged high achievers, the rational expectation of an hu receiving financial aid to attend such a university is that he will not derive the full benefit of contact with hp's. Insofar as there can be segregation within the university, this hu student may prefer a second-tier university ((hu, hu) or $(\ell p, hu)$) instead.

The options for the universities are either to impose more aggressive association policies or to find ways of replicating full transferability. As we have noted, the military academies apart, colleges have been reluctant to go beyond certain random association policies (the ones we have examined may be more aggressive than those used in practice).[16] As for improving transferability, our analysis shows that one needs to mimic the provision of side payments *to the privileged*. From this perspective, financial aid through scholarships awarded only on the basis of "need" may undermine diversity goals. What an hp requires in order to associate with an hu is a compensating transfer; one way to implement that would be to offer hp's scholarships contingent on some degree of desired associational behavior.[17]

We have focused on the composition of peer groups within colleges as the source of excessive segregation. Making this the only benefit of college is, of course, a simplification. The usual focus of diversity policies is at the admission level, and given our payoff structure, these policies by themselves will have no effect. By contrast, if part of the college premium is due to quality of faculties and facilities, policies targeting only admission may have some bite. For instance, if there are complementarities between faculty and students, a college admitting only high achievers, but on a color blind basis, will benefit the hu's at the expense of the hp's, although there is still segregation within the college. As long as the complementarities are not too strong (a modified condition (DD) holds), the other results would be similar: the free market outcome will be segregation under NTU, with inequality between backgrounds exacerbated by the difference in faculty qualities. In the first best, different attributes (hp and hu) would interact; the free market

[16]Firms, unlike universities, have little reluctance to exercise managerial authority in the assignment of employees into teams or work groups and regularly use tools such work station restrictions or reporting structures that can regulate association. Thus, whatever the relative interest of firms and universities in achieving diversity, firms arguably have a more powerful array of instruments to get there.

[17]If this sounds implausibly interventionist, observe that in other arenas, U.S. colleges have been more than willing to manage the transactional behavior of athletes, particularly those with scholarships.

under NTU will lead to distorted investments, and policies that affect both admission and association will improve matters more than admission alone.

It is not obvious, however, that faculty quality is necessarily a complement to student quality, since worse students may benefit more from better teaching. In this case policy analysis incorporating college heterogeneity will be more subtle, pointing to a promising direction for future research.

Another question concerns relaxing the assumption that both backgrounds have the same investment costs. It is straightforward to modify the model to allow, for example, higher marginal costs for the underprivileged. This will tend to mitigate the benefits of an affirmative action policy, both because the underprivileged's investments will be less responsive, and because the privileged, now less likely to interact with the underprivileged, will reduce their investment less. A pertinent observation is that investments often happen in environments such as school or neighborhoods, in which there are peer effects and in which the market outcome is characterized by similar imperfections as the one we considered here. Re-matching policies can be applied at the school or neighborhood level as well as at college, and this raises questions of how re-matching policies in one level impact on the performance of policies in another, as well as the complementarity or substitutability of re-matching policies in sequential markets.

A Appendix: Proofs

A.1 Proofs for Section 3

Proof of Lemma 1

Using (i)-(iii), and reversing the argument (iv), noting that under $\delta > \alpha + \beta$ $(hu, \ell p)$ matches induce higher surplus than the sum of partners' segregation payoffs, the possible stable heterogeneous peer groups are (hp, hu), $(hu, \ell p)$, and $(\ell p, \ell u)$ (i.e., all three matches will be formed if the alternative is segregation). Reversing the argument in (v), if $\alpha > 1-\delta$ having matches (hp, hu) and segregating ℓp induces higher surplus than $(hu, \ell p)$ matches and segregating hp students. Hence, under the condition, (hp, hu) matches are exhausted. Comparing matches $(hu, \ell p)$ and segregating ℓu students, yielding surplus $\delta + \alpha\beta$ to matching $(\ell p, \ell u)$ and segregating hu students, yielding surplus $\alpha\delta + \beta$, the former surplus is higher than the latter if $\delta > \beta$, as assumed.

A.2 Proofs for Section 4

Proof of Proposition 2

We show that either policy on its own leads to segregation.

Suppose first there is an admission policy in place, so that $q_c(ab) > 0$ and $q_c(a'b') > 0$ for some college c. If this is part of a college market equilibrium, then $p_c(ab, ab) = 1$ and $p_c(a'b', a'b') = 1$. Suppose otherwise. Let ab induce higher payoff $y(ab, .)$ to any other attribute than does $a'b'$ wlog. Since association is free within college, then ab students have strictly higher payoff if the college c sets $p_c(ab, ab) = 1$ instead, which violates the stability condition of the college market equilibrium. This argument extends by induction to all other attributes $a'b'$ with $q_c(a'b') > 0$, starting with the attribute that induces the second highest payoff after ab, and so on.

Now suppose an association policy is in place, that is colleges use $p_c(ab, a'b') \in (0, 1)$ for some combination of $ab \neq a'b'$. Suppose that ab induces higher payoff $y(ab, .)$ to any other attribute than does $a'b'$. An allocation that entails $q_c(ab) > 0$ and $q_c(a'b') > 0$ is not a college market equilibrium. Suppose otherwise. Then there is an admissible college c' setting $q_{c'}(ab) = 1$ and $u(ab|c') = y(ab, ab) > p_c(ab, ab)y(ab, ab) + \sum_{a'b' \neq ab} p_c(ab, a'b')y(ab, a'b') = u(ab|c)$, and college c gave highest payoff to both ab and $a'b'$ by the assumption that $q_c(ab) > 0$ and $q_c(a'b') > 0$ in equilibrium.

Proof of Lemma 4

Let $Y \equiv [y(ab, a'b')]$ be the 4×4 matrix of individual payoffs under NTU, and Y_c the sub-matrix obtained by deleting the rows and columns of Y that are not in the support of c. Consider two colleges with the same support of types; since there is random matching, for each pair of types, $p_c(ab, a'b') = q_c(a'b')$. Let q_c be the vector consisting of the positive probabilities $q_c(a'b')$; the dimension of q_c is equal to the dimension of the square matrix Y_c, and the expected payoffs of the different types in the support of c are given by the product $q_c' \cdot Y_c$. Since each type in the support of c, c' must be indifferent between the two colleges, we must have $(q_c' - q_{c'}') \cdot Y_c = 0$, or $q_c' - q_{c'}'$ must belong to the null-space of Y_c. The result follows if, and only if, the null space coincides with $\{0\}$ for almost all parameter values, or alternatively if the rows of Y_c are linearly independent, something we show in Lemma 9 in the Appendix.

Proof that the Kernel of Y_c is equal to 0.

Lemma 9. *For any college c, if Y_c is the matrix of payoffs of types in the support of c, then Y_c has a kernel equal to $\{0\}$ for a generic set of parameters.*

If a matrix is obtained from making a linear transform of a row (column) or adding such a linear transform to another row (column), the two matrices have the same null-space.

Suppose first that a college has full support, then (the columns and rows are ordered by $hp, \ell p, hu, \ell u$)

$$Y_c = \begin{bmatrix} 1 & \delta & 1/2 & \delta/2 \\ \delta & \beta & \delta/2 & \beta/2 \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix}$$

If, for instance, the second row is modified by substracting a multiple m of the fourth row, we denote the resulting change in the matrix as $\xrightarrow{r2-mr4}$. We have:

$$\begin{aligned} Y &\xrightarrow{r2-2r4} \begin{bmatrix} 1 & \delta & 1/2 & \delta/2 \\ 0 & 0 & \delta(1/2-2\alpha) & \beta(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix} \\ &\xrightarrow{r1-2r3} \begin{bmatrix} 0 & 0 & 1/2-2\alpha & \delta(1/2-2\alpha) \\ 0 & 0 & \delta(1/2-2\alpha) & \beta(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix} \\ &\xrightarrow{r2-\delta r1} \begin{bmatrix} 0 & 0 & 1/2-2\alpha & \delta(1/2-2\alpha) \\ 0 & 0 & 0 & (\beta-\delta)(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix} \\ &\xrightarrow{r4-\delta r3} \begin{bmatrix} 0 & 0 & 1/2-2\alpha & \delta(1/2-2\alpha) \\ 0 & 0 & 0 & (\beta-\delta)(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ 0 & (\beta-\delta^2)/2 & 0 & \alpha(\beta-\delta^2) \end{bmatrix} \equiv \hat{Y} \end{aligned}$$

Solving $x'\hat{Y} = 0$, the first equation is $x_3/2 = 0$ which implies $x_3 = 0$. As long

as $\beta \neq \delta^2$, $\alpha \neq 1/4$ and $\beta \neq \delta$, in the second equation $\delta x_3/2 + (\beta - \delta^2)x_4/2$ we must have $x_4 = 0$; the third equation then implies that $x_1 = 0$, and the last equation that $x_2 = 0$. Hence the null space of Y is $\{0\}$.

Similar reasoning can be made for all sub-matrices obtained from Y by removing the row and column of types which are not in the support of c. For instance if college c has support $\{hu, \ell p\}$,

$$Y_c = \begin{bmatrix} \beta & \delta/2 \\ \delta/2 & \alpha \end{bmatrix} \xrightarrow{r1-2\alpha\beta r2/\delta} \hat{Y}_c \equiv \begin{bmatrix} 0 & \delta/2 - 2\alpha/\delta \\ \delta/2 & \alpha \end{bmatrix}$$

the first equation in $x' \cdot \hat{Y}_c = 0$ is $x_2\delta/2 = 0$ implies $x_2 = 0$; the second equation equation then implies $x_1 = 0$ whenever $\delta^2 \neq 4\alpha$.

Proof of Proposition 5

In the proof of Proposition 4 we showed that $W^B < W^0$ for all π. The second and third parts of the proposition are proved as part of the proof of Proposition 3.

A.3 Proofs for Section 5

Proof of Lemma 7

In a college market equilibrium hp and ℓu students must segregate, i.e. $p_c(\ell u, \ell u) = 1$ and $p_c(hp, hp) = 1$ with tuition fees $t_c(\ell u, \ell u) = t_c(hp, hp) = 0$ for all colleges c. This is because hp students cannot be adequately compensated by any other attribute and ℓu cannot adequately compensate any other attribute. hu and ℓp agents cannot both segregate ($p_c(hp, \ell u) = 0$), since a transfer from ℓp to hu of $t(\ell p, hu) = \beta - \delta/2 + 2\epsilon$ and $t(hu, \ell p) = -\beta + \delta/2 - \epsilon$ would make both sides strictly better off. Hence, within a college $t(hu, \ell p) = \beta - \delta/2$ and $p_c(hp, \ell p) < 1$ if $q_c(hu) > q_c(\ell p)$, $t(hu, \ell p) = \delta - \beta$ and $p_c(hp, \ell p) = 1$ if $q_c(hp) < q_c(\ell p)$, and $t(hu, \ell p) \in [\beta - \delta/2, \delta/2 - \alpha]$ and $p_c(hp, \ell p) = 1$ if $q_c(hp) = q_c(\ell p)$.

Thus in a college market equilibrium stability implies that hu and ℓp students will not segregate as there is a transfer from ℓu to hu that makes both strictly better off. Moreover, $p_c(hu, \ell p) = p_{c'}(hu, \ell p)$ implies $t_c(hu, \ell p) = t_{c'}(hu, \ell p)$. Therefore there cannot be two colleges with $q_c(hu) > q_c(\ell p)$ and $q_{c'}(hu) < q_{c'}(\ell p)$, since hu students would strictly profit from switching from

c to c' obtaining higher transfers and less interaction with ℓp. Hence, all colleges with ℓp and hu students will have $p_c(hu, \ell p) = \frac{\pi(1-e_p)}{(1-\pi)e_u+\pi(1-e_p)}$.

Proof of Lemma 8

Note first that still hp students cannot be compensated by a side payment from any ℓ student. Hence, $p_c(hp, \ell b) = 0$ in all colleges with $q_c(hp) > 0$.

ℓu students cannot compensate any other attribute for their negative local externalities since they have not enough wealth. Therefore $p_c(\ell u, \ell u) = 1$ in every college with $q_c(\ell u) > 0$ and ℓu segregate into ℓu colleges.

Since hu have priority over hp students, they have the choice between a college with $p_c(hu, hp) > 0$ and zero transfers and colleges with $p_c(hu, \ell p) > 0$, but receiving a transfer. For an hu to be indifferent:

$$p_c(hu, hp)(\delta - \beta) = p_{c'}(hu, \ell p)(1/2 + t_{c'}(hu, \ell p) - \beta).$$

Since $t(hu, \ell p) \leq \delta/2 - \alpha$, because otherwise ℓp would prefer to segregate, for high $p_c(hu, hp)$ also ℓp segregate. This is the case if hp are abundant compared to hu, since no arbitrage implies that college composition reflects the population shares of hu and hp. For smaller population shares of hp, hu will be made indifferent by an equilibrium transfer between colleges with hp and hu and those with hu and ℓp.

Denote by ρ the share of hu students who attend (hu, hp) universities under an $\overline{A}$ policy. To make hu students indifferent between both types of colleges the transfer has to satisfy:

$$\frac{\pi e_p}{(1-\pi)e_u\rho + \pi e_p}(\delta - \beta) = \frac{\pi(1-e_p)}{(1-\pi)e_u(1-\rho) + \pi(1-e_p)}(\delta/2 + t(hu, \ell p) - \beta).$$

The share ρ implicitly defined above decreases in $t(hu, \ell p)$, and $\rho > 1$ for $t(hu, \ell p) = \beta - \delta/2$. In order to have colleges with both hu and ℓp, $v(\ell p) \geq \alpha$, that is $t(hu, \ell p) \leq \delta/2 - \alpha$. Therefore, for colleges with both ℓp and hu to form, $\rho \leq 1$ for $t(hu, \ell p) \leq \delta/2 - \alpha$, which yields:

$$\frac{\pi e_p}{\pi(1-e_p)} \frac{(1-\pi)e_u(1-\rho) + \pi(1-e_p)}{(1-\pi)e_u\rho + \pi e_p}(\delta - \beta) \leq \delta - \beta - \alpha,$$

for $\rho \leq 1$. Since the left hand decreases in ρ colleges with both ℓp and hu

form if and only if:

$$\pi < \frac{(\delta - \beta - \alpha)e_u^A}{(\delta - \beta - \alpha)e_u^A + \alpha e_p^A} := \pi^*.$$

B Online Appendix: Computations

B.1 Proofs for Section 3

Proof of Lemma 2

Depending on relative scarcity of hu, ℓp, and hp agents there are five cases.

Case (1): $\pi e_p > (1-\pi)e_u$ and $\pi(1-e_p) > (1-\pi)(1-e_u)$: Then some hp segregate and $v(hp) = 1$. hu match with hp and obtain $v(hu) = 2\delta - 1$. Likewise, some ℓp remain unmatched and obtain $v(\ell p) = \alpha$, whereas $v(\ell u) = (2\delta - 1)\alpha$. Hence, $e_p = 1 - \alpha$ and $e_u = (2\delta - 1)(1 - \alpha)$. The conditions become

$$\frac{\pi}{1-\pi} > \max\{2\delta - 1; (1 - (1-\alpha)(2\delta - 1))/\alpha\} = \frac{1 - (1-\alpha)(2\delta - 1)}{\alpha}.$$

Case (2): $\pi e_p > (1-\pi)e_u$ and $\pi(1-e_p) < (1-\pi)(1-e_u)$: Then $v(hp) = 1$ and $v(hu) = 2\delta - 1$ as above. But now $v(\ell u) = \alpha\beta$ and $v(\ell p) = \alpha(2\delta - \beta)$. Hence, $e_p = 1 - \alpha(2\delta - \beta)$ and $e_u = 2\delta - 1 - \alpha\beta$. The conditions become

$$\frac{2\delta - 1 - \alpha\beta}{1 - \alpha(2\delta - \beta)} < \frac{\pi}{1-\pi} < \frac{2 - 2\delta + \alpha\beta}{\alpha(2\delta - \beta)}.$$

Case (3): $\pi e_p < (1-\pi)e_u$ and $\pi > 1-\pi$. Then some ℓp segregate, so that $v(\ell p) = \alpha$. Therefore $v(hu) = \delta - \alpha$ and $v(hp) = \delta + \alpha$. $v(\ell u) = \alpha(2\delta - 1)$. Therefore $e_p = \delta$ and $e_u = (1 - 2\alpha)\delta$. The first condition then would imply $\pi/(1-\pi) < 1 - 2\alpha$, which is a contradiction to the second, $\pi/(1-\pi) > 1$.

Case (4): $\pi e_p < (1-\pi)e_u < \pi$ and $\pi < 1 - \pi$. Now some ℓu segregate, so that $v(\ell u) = \alpha\beta$. Therefore $v(\ell p) = \alpha(2\delta - \beta)$ and $v(hu) = \delta - \alpha(2\delta - \beta)$ and $v(hp) = \delta + \alpha(2\delta - \beta)$. This means that $e_p = \delta$ and $e_u = (1 - 2\alpha)\delta$. The conditions become

$$(1 - 2\alpha)\delta < \frac{\pi}{1-\pi} < 1 - 2\alpha.$$

Case (5): $\pi < (1-\pi)e_u$: Now some hu segregate, so that $v(hu) = \beta$ and $v(\ell u) = \alpha\beta$. $v(hp) = 2\delta - \beta$ and $v(\ell p) = \delta - \beta$, so that $e_p = \delta$ and

$e_u = (1-\alpha)\beta$. The condition becomes

$$\frac{\pi}{1-\pi} < (1-\alpha)\beta.$$

The intermediate cases where e_p and e_u are determined by $\pi(1-e_p) = (1-\pi)(1-e_u)$, $\pi e_p = (1-\pi)e_u < \pi$, and $e_u = \pi/(1-\pi)$ are omitted. To summarize, for

- $\pi \leq \frac{1-2\alpha}{2(1-\alpha)}$, $e_p = \delta$.
- $\frac{1-2\alpha}{2(1-\alpha)} < \pi < \frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)}$ e_p strictly decreases,
- $\frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)} \leq \pi \leq \frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)}$ e_p reaches a minimum at $e_p = 1-\alpha(2\delta-\beta)$.
- $\frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)} < \pi < \frac{2(1-\delta(1-\alpha))-\alpha}{2(1-\delta(1-\alpha))}$ e_p strictly increases.
- $\pi \geq \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)} \equiv \hat{\pi}_2$, $e_p^* = 1-\alpha$.

Similarly, for

- $\pi \leq \frac{(1-\alpha)\beta}{1+(1-\alpha)\beta}$, $e_u = (1-\alpha)\beta$.
- $\hat{\pi}_0 \equiv \frac{(1-\alpha)\beta}{1+(1-\alpha)\beta} < \pi < \frac{(1-2\alpha)\delta}{1-(1-2\alpha)\delta}$ e_u strictly increases,
- $\frac{(1-2\alpha)\delta}{1-(1-2\alpha)\delta} \leq \pi \leq \frac{1-2\alpha}{2-2\alpha}$, $e_u = (1-2\alpha)\delta$,
- $\frac{1-2\alpha}{2-2\alpha} < \pi < \frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)}$ e_u strictly increases,
- $\frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)} \leq \pi \leq \frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)}$ e_u reaches a maximum at $e_u = 2\delta-1-\alpha\beta$,
- $\frac{2(1-\delta)+\alpha\beta}{2(1-\delta+\alpha\delta)} < \pi < \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_u = 1-\alpha)(2\delta-1)$ e_u strictly decreases.
- $\pi \geq \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_u = (1-\alpha)(2\delta-1)$.

Let $\hat{\pi}_2$ be defined by $e_p = 1-\alpha$, that is by

$$\hat{\pi}_1 = \frac{1-2\alpha}{2(1-\alpha)} + \left(\frac{2\delta-1-\alpha\beta}{2\delta(1-\alpha)} - \frac{1-2\alpha}{2(1-\alpha)}\right)\frac{\delta-1+\alpha}{\delta-(1-\alpha(2\delta-\beta))}.$$

Proof of Lemma 3

To establish static surplus efficiency, suppose the contrary, i.e., a set of agents can be rematched to increase total payoff of all these agents. Then the increase in total payoff can be distributed among all agents required to rematch, which makes all agents required to re-match also strictly prefer their new

matches, a contradiction to stability. Therefore matching is surplus efficient given investments.

The second part of the lemma requires some work. Let $\{ab\}$ denote a distribution of attributes in the economy, and $\mu(ab, a'b')$ the measure of $(ab, a'b')$ matches in a surplus efficient match given $\{ab\}$. Since $\mu(ab, a'b')$ only depends on aggregates πe_p, $\pi(1-e_p)$, $(1-\pi)e_u$, and $(1-\pi)(1-e_u)$ and investment cost is strictly convex, in an allocation maximizing total surplus all p agents invest the same level e_p, and all u agents invest e_u.

An investment profile (e_u, e_p) and the associated surplus efficient match $\mu(.)$ maximize total surplus ex ante if there is no (e'_u, e'_p) and an associated surplus efficient match $\mu(.)$ such that total surplus is higher.

Denote the change in total surplus Δ_b by increasing e_b to $e'_b = e'_b + \epsilon$. If there are positive measures of (hp, hp) and (hp, hu) schools, it is given by:

$$\Delta_p = \epsilon[z(hp, hu) - z(\ell p, hu)] - \epsilon e_p - \epsilon^2/2 \text{ and}$$
$$\Delta_u = \epsilon[z(hp, hu) - z(hp, hp)/2] - \epsilon e_u - \epsilon^2/2,$$

reflecting the gains from turning an ℓp student matched to an hu student into an hp student matched to an hu, and from turning an ℓu student matched to an ℓu student into an hu student matched to an hp, who used to be matched to an hp.

That is, assuming that indeed $\pi > (1-\pi)e_u > \pi(1-e_p)$ the optimal investments are given by $e_p = z(hp, hp)/2$ and $e_u = z(hp, hu) - z(hp, hp)/2$. Recall that TU wages are given in this case by $v(hp) = z(hp, hp)/2 = 1$ and $v(\ell p) = z(hu, \ell p) - v(hu)$, and $v(hu) = z(hp, hu) - z(hp, hp)/2 = 2\delta - 1$ and $y(\ell u) = 0$. Hence, TU investments are $e_p^T = z(hp, hu) - z(hu, \ell p)$ and $e_u^T = z(hp, hu) - z(hp, hp)/2$. That is, TU investments are optimal with respect to marginal deviations.

To check for larger deviations suppose only e_u increases by ϵ, such that the measure of (hu, hu) firms becomes positive after the increase. The change in total surplus is now:

$$\Delta = \epsilon_1[z(hp, hu) - z(\ell p, hu)] + \epsilon_2[z(hu, hu)/2 - z(\ell u, \ell u)/2] - \epsilon e_p - \epsilon^2/2,$$

for $\epsilon_1 + \epsilon_2 = \epsilon$ such that the measure of (hp, hp) under e_u was $\epsilon_1/2$. Clearly, $\Delta < 0$ for $e_u = z(hp, hu) - z(\ell p, hu)$, since cost is convex and surplus has decreasing returns in an efficient matching. Suppose now that e_p decreases

by ϵ large enough to have a positive measure of $(\ell p, \ell p)$ students after the decrease (a decrease in e_u would have the same effect). The change in total surplus is:

$$\Delta = -\epsilon_1[z(hp, hu) - z(\ell p, hu)] - \epsilon_2[z(hp, hp)/2 - z(\ell p, \ell p)/2] + \epsilon e_p - \epsilon^2/2,$$

which is negative for $e_p = z(hp, hu) - z(hu, \ell p)$ since cost is convex and surplus has decreasing returns in an efficient matching. Finally, an increase of e_p will not affect the condition $\pi > (1-\pi)e_u > \pi(1-e_p)$.

A similar argument holds in all the five cases present in the proof of Fact 2.

B.2 Proofs for Section 4

Proof of Proposition 3

Starting with an $\overline{A}$ policy, the payoffs resulting from the lemma are $v(\ell u) = \alpha\beta$, $v(\ell p) = \alpha$, and

$$v(hu) = \frac{\pi e_p \delta + (1-\pi)e_u\beta}{\pi e_p + (1-\pi)e_u} \text{ and } v(hp) = \frac{\pi e_p + (1-\pi)e_u\delta}{\pi e_p + (1-\pi)e_u}.$$

Optimal investments anticipating the equilibrium measures are therefore

$$e_u = \frac{\pi e_p(\delta - \beta)}{\pi e_p + (1-\pi)e_u} + \beta(1-\alpha) > e_u^0, \tag{B.1}$$

and

$$e_p = 1 - \alpha - \frac{(1-\pi)e_u(1-\delta)}{\pi e_p + (1-\pi)e_u} < e_p^0. \tag{B.2}$$

Rewriting and dividing the second by the first equation yields:

$$e_p = \frac{1-\delta}{\delta - \beta}e_u + \delta - \alpha - \frac{1-\delta}{\delta-\beta}\beta(1-\alpha). \tag{B.3}$$

This implies that e_p increases in e_u at a rate of less than unity. Using this fact and the above expression on (B.1) reveals that both e_p and e_u must increase

in π, and yields a quadratic expression for e_u:

$$\begin{aligned}0 =& e_u^2\left(\frac{1-\pi}{\pi}+\frac{1-\delta}{\delta-\beta}\right)+e_u\left(\delta-\alpha-\frac{1-\delta}{\delta-\beta}[\delta+\beta-2\alpha\beta]-\frac{1-\pi}{\pi}\beta(1-\alpha)\right)\\ &-(\delta-\alpha\beta)(\delta-\alpha-\frac{1-\delta}{\delta-\beta}\beta(1-\alpha)).\end{aligned}$$

For future reference the differential of e_u and π is:

$$\frac{\partial e_u}{\partial \pi}=\frac{(\delta-\alpha\beta-e_u)e_p+(e_u-(1-\alpha)\beta)e_u}{\pi(e_p+(\delta-\alpha\beta-e_u)\frac{1-\delta}{\delta-\beta})+(1-\pi)(2e_u-(1-\alpha)\beta)}>0. \qquad \text{(B.4)}$$

Lower investment inequality (i.e., $e_p/e_u < 1/\beta = e_p^0/e_u^0$) follows directly from the expressions for e_p and e_u above. Notice that $e_u > e_p$ for $\pi = 1$ if $1-\delta < \alpha(1-\beta)$, which is possible under our assumptions. Because of continuity the second part of that statement follows. Payoff inequality, given by $\frac{e_p^2+v(\ell p)}{e_u^2+v(\ell u)}$ must be greater under the free market, because $v^0(\ell b) = v(\ell b)$ for $b = u, p$, and both $e_p < e_p^0$ and $e_u > e_u^0$.

For the remaining assertions start with aggregate investment. It is higher under the $\overline{A}$ policy than under laissez faire if

$$\pi e_p+(1-\pi)e_u>\pi(1-\alpha)+(1-\pi)\beta(1-\alpha).$$

Using the expressions above this becomes

$$-\pi\frac{(1-\pi)e_u(1-\delta)}{\pi e_p+(1-\pi)e_u}+(1-\pi)\frac{\pi e_p(\delta-\beta)}{\pi e_p+(1-\pi)e_u}>0.$$

For $0 < \pi < 1$ this simplifies to

$$e_p(\delta-\beta)>e_u(1-\delta).$$

Using (B.3) we have:

$$\frac{\delta-\beta}{1-\delta}(\delta-\alpha)>\beta(1-\alpha).$$

Under our assumptions ($1-\delta < \alpha < \delta-\beta$) this must be true.

For aggregate output Y in the economy (that is, aggregate production in

matches net of effort cost) and aggregate welfare W notice that generally:

$$W = \pi\frac{e_p^2}{2} + \pi v(\ell p) + (1-\pi)\frac{e_u^2}{2} + (1-\pi)v(\ell u),$$

and

$$Y = \pi e_p^2 + \pi v(\ell p) + (1-\pi)e_u^2 + (1-\pi)v(\ell u).$$

Since under an $\overline{A}$ policy $v(\ell u) = v^0(\ell u)$ and $v(\ell p) = v^0(\ell p)$ the welfare comparison reduces to:

$$W^{\overline{A}} - W^0 = \pi\frac{(e_p)^2 - (e_p^0)^2}{2} + (1-\pi)\frac{(e_u)^2 - (e_u^0)^2}{2},$$

and $W^{\overline{A}} > W^0 \Leftrightarrow Y^{\overline{A}} > Y^0$. That is, $W^{\overline{A}} > W^0$ if

$$(1-\pi)(e_u - e_u^0)(e_u + e_u^0) > \pi(e_p^0 - e_p)(e_p^0 + e_p). \tag{B.5}$$

Using the expressions for e_u and e_p from above:

$$\frac{e_u - e_u^0}{e_p^0 - e_p} = \frac{\pi e_p(\delta - \beta)}{(1-\pi)e_u(1-\delta)}.$$

Using this expression on (B.5) yields

$$\frac{e_p}{e_u}\frac{\delta - \beta}{1-\delta} > \frac{e_p^0 + e_p}{e_u^0 + e_u}.$$

From above we know that $e_p \leq 1 - \alpha$ and $e_u \geq \beta(1-\alpha)$ so that the above expression is satisfied if

$$\beta\frac{e_p}{e_u}\frac{\delta - \beta}{1-\delta} > 1.$$

Using (B.3), the ratio $\frac{e_p}{e_u}$ decreases in e_u and also in π. This means this ratio is bounded below by $(1-\alpha)/(\delta - \alpha\beta))$, and that a sufficient condition for $W^{\overline{A}} > W^0$ for all $\pi \in (0,1)$ is:

$$\beta(\delta - \beta)(1-\alpha) > (1-\delta)(\delta - \alpha\beta).$$

This condition is satisfied for δ sufficiently close to 1, or if $(\delta - \beta) - (1-\delta)$ sufficiently great. Hence, there is $\hat{\delta} < 1$ such that for $\delta > \hat{\delta}$ both aggregate surplus W and aggregate payoffs Y are higher under the $\overline{A}$ policy.

Suppose now that an $\overline{A}$ policy is place yielding e_p and e_u. Compare

this outcome to one that arises from a modified A policy, an A' policy that assigns probability $\epsilon > 0$ on ℓp students to match with ℓu students. Denote the corresponding outcome by e_p^ϵ and e_u^ϵ. Under the perturbation $v(\ell p) = v^{\overline{A}}(\ell p) - \epsilon\alpha(1-\delta) < v^{\overline{A}}(\ell p)$, and, because of measure consistency, $v(\ell u) = v^{\overline{A}}(\ell u) + \frac{\pi}{1-\pi}\frac{1-e_p^\epsilon}{1-e_u^\epsilon}\epsilon\alpha(\delta-\beta) > v^{\overline{A}}(\ell u)$. As above:

$$v(hu) = \frac{\pi e_p^\epsilon \delta + (1-\pi)e_u^\epsilon \beta}{\pi e_p^\epsilon + (1-\pi)e_u^\epsilon} \text{ and } v(hp) = \frac{\pi e_p^\epsilon + (1-\pi)e_u^\epsilon \delta}{\pi e_p^\epsilon + (1-\pi)e_u^\epsilon}.$$

Since $e_u^\epsilon = v(hu) - v(\ell u)$ and $e_p^\epsilon = v(hp) - v(\ell p)$, and $v(\ell p)$ decreases and $v(\ell u)$ increases in ϵ, and both $v(hu)$ and $v(hp)$ increase in e_u^ϵ and decrease in e_p^ϵ, it must be the case that $v(hu)$ and $v(hp)$ both increase in ϵ. Intuitively, the quality of the pool of h students has improved.

Aggregate surplus is given by

$$\pi\frac{(v(hp)-v(\ell p))^2}{2} + \pi v(\ell p) + (1-\pi)\frac{(v(hu)-v(\ell u))^2}{2} + (1-\pi)v(\ell u).$$

Taking the first derivative with respect to ϵ yields:

$$\pi e_p^\epsilon \frac{\partial v(hp)}{\partial \epsilon} - \pi(1-e_p^\epsilon)\alpha(1-\delta) + (1-\pi)e_u^\epsilon \frac{\partial v(hu)}{\partial \epsilon} + (1-\pi)(1-e_u^\epsilon)\frac{\pi}{1-\pi}\frac{1-e_p^\epsilon}{1-e_u^\epsilon}\alpha(\delta-\beta).$$

This simplifies to

$$\pi e_p^\epsilon \frac{\partial v(hp)}{\partial \epsilon} + (1-\pi)e_u^\epsilon \frac{\partial v(hu)}{\partial \epsilon} + \pi(1-e_p^\epsilon)\alpha(2\delta-\beta-1) > 0,$$

which follows from the assumption $1-\delta < \delta-\beta$ and the fact that $\frac{\partial v(hb)}{\partial \epsilon} > 0$ for both $b = u, p$. Note that the argument extends to initial configurations with $\epsilon > 0$. Hence, an increase in ϵ, conditional on the resulting college distributions being measure consistent, will increase aggregate surplus. Therefore, an A' policy awarding priority for u applicants over p applicants with the same achievement will achieve higher aggregate surplus than an $\overline{A}$ policy, and, in particular, an A policy.

An analogous argument can be used to establish that aggregate output, i.e. $\pi(e_p^\epsilon v(hp) + (1-e_p^\epsilon)v(\ell p)) + (1-\pi)(e_u^\epsilon v(hu) + (1-e_u^\epsilon)v(\ell u))$, is higher under any modified A' policy, at least if $e_p^\epsilon \geq e_u^\epsilon$, which must be the case for an A policy as shown below.

Moreover, since $v(hp) \leq 1$ and $v(\ell p) < \alpha$ for $\pi < 1$ it must be the case

that under any modified A policy $e_p^\epsilon < e_p^0$. Similarly, under any modified A policy $e_u^\epsilon = v(hu) - v(\ell u) > \beta(1-\alpha)$ if $(1-\pi)\frac{e_p^\epsilon}{\pi e_p^\epsilon + (1-\pi)e_u^\epsilon} > \frac{1-e_p^\epsilon}{1-e_u^\epsilon}\alpha\epsilon$, which must be true if $e_p^\epsilon > e_u^\epsilon$ (and otherwise $e_p^\epsilon/e_u^\epsilon < e_p^0/e_u^0$ trivially), and thus investment inequality is lower under the modified A policy than in the free market. Higher investment inequality and the fact that $v(\ell u) > \alpha\beta$ and $v(\ell p) < \alpha$ for $0 < \pi < 1$ under the modified A policy imply that surplus inequality $\frac{(e_p^\epsilon)^2/2 + v(\ell p)}{(e_u^\epsilon)^2/2 + v(\ell u)}$ is smaller than in the free market.

Aggregate investment under a modified A policy is higher than in the free market if:

$$\pi e_p^\epsilon + (1-\pi)e_u^\epsilon > \pi(1-\alpha) + (1-\pi)\beta(1-\alpha),$$

Under an A policy, matching ℓu and ℓp uniformly to each other the two investment levels are proportionate (which can be derived by manipulating the expressions for e_u and e_p as in the case of an A policy above):

$$e_p = \frac{1-\delta}{\delta-\beta}e_u + \frac{2\delta - 1 - \beta}{\delta - \beta} > e_u.$$

Using this expression aggregate investment is higher under the A policy if:

$$\pi\left(\frac{2\delta-1-\beta}{\delta-\beta}(1-e_u) - (1-\alpha)(1-\beta)\right) + e_u - \beta(1-\alpha) > 0.$$

The LHS of this condition strictly increases in $e_u \geq (1-\alpha)\beta$, so that a sufficient condition is given by

$$\frac{2\delta-1-\beta}{\delta-\beta}(1-\beta(1-\alpha)) > (1-\alpha)(1-\beta),$$

which yields, after some manipulation:

$$\frac{\alpha}{1-\beta(1-\alpha)} > \frac{1-\delta}{\delta-\beta}.$$

This condition must hold for δ close enough to 1.

Proof of Proposition 4

Under a B policy students are distributed across colleges according to the population measures. Therefore individual payoffs are given as:

$$\begin{aligned}
v(hp) &= \pi e_p + (1-\pi)e_u\delta + \pi(1-e_p)/2 + (1-\pi)(1-e_u)\delta/2,\\
v(\ell p) &= \pi e_p/2 + (1-\pi)e_u\delta/2 + \pi(1-e_p)\alpha + (1-\pi)(1-e_u)\alpha\delta,\\
v(hu) &= \pi e_p\delta + (1-\pi)e_u\beta + \pi(1-e_p)\delta/2 + (1-\pi)(1-e_u)\beta/2,\\
v(\ell u) &= \pi e_p\delta/2 + (1-\pi)e_u\beta/2 + \pi(1-e_p)\alpha\delta + (1-\pi)(1-e_u)\alpha\beta.
\end{aligned}$$

This implies investment choices satisfy:

$$\begin{aligned}
e_p &= \pi(1/2-\alpha) + (1-\pi)(1/2-\alpha)\delta + \pi e_p\alpha + (1-\pi)e_u\alpha\delta,\\
e_u &= \pi(1/2-\alpha)\delta + (1-\pi)(1/2-\alpha)\beta + \pi e_p\alpha\delta + (1-\pi)e_u\alpha\beta.
\end{aligned}$$

Using the expressions in the text, optimal investments under the B policy are given by:

$$\begin{aligned}
e_p^B &= (1/2-\alpha)\frac{\pi + (1-\pi)\delta + \pi(1-\pi)\alpha(\delta^2-\beta)}{\pi(1-\alpha) + (1-\pi)(1-\alpha\beta) - \pi(1-\pi)\alpha^2(\delta^2-\beta)},\\
e_u^B &= (1/2-\alpha)\frac{\pi\delta + (1-\pi)\beta + \pi(1-\pi)\alpha(\delta^2-\beta)}{\pi(1-\alpha) + (1-\pi)(1-\alpha\beta) - \pi(1-\pi)\alpha^2(\delta^2-\beta)}.
\end{aligned}$$

This immediately implies that $e_p^B/e_u^B < 1/\beta = e_p^0/e_p^0$. Since payoffs are given by $y_u^B = (e_u^B)^2 + v^B(\ell u)$ and $y_u^0 = (e_u^0)^2 + \alpha\beta$ and analogously for p students, $e_p^B/e_u^B < 1/\beta = e_p^0/e_p^0$ and $e_p^B < e_p^0$ and $v^B(\ell p) < \beta v^(\ell u)$ and $v^B(\ell u) > \alpha\beta$ also imply that $\frac{(e_p^B)^2+v^B(\ell p)}{(e_u^B)^2+v^B(\ell u)} < \frac{(1-\alpha)^2+\alpha}{\beta^2(1-\alpha)^2+\alpha\beta}$. Therefore $y_p^B/y_u^B < y_p^0/y_u^0$.

It is quickly verified by differentiation that both e_u^B and e_p^B increase in π. Therefore $e_p^B \le (1/2-\alpha)/(1-\alpha) < \delta/2 < (1-\alpha) = e_p^0$, and $e_p^B < \delta/2 < \delta - \alpha \le e_p^A$ using that $1-\delta < \alpha < \delta/2$. For the underprivileged $e_u^B \le \delta(1/2-\alpha)/(1-\alpha) < \beta(1-\alpha) = e_u^0 < e_u^A$. Therefore aggregate investments are smaller under the B policy: $\pi e_p^B + (1-\pi)e_u^B < \pi e_p^0 + (1-\pi)e_u^0$. Moreover, $e_u^B < e_p^B < 1/2$ and the FOCs above imply that

$$\begin{aligned}
e_p^B &< \frac{1-\alpha}{2}(\pi + (1-\pi)\delta),\\
e_u^B &< \frac{1-\alpha}{2}(\pi\delta + (1-\pi)\beta).
\end{aligned}$$

Aggregate welfare under a B policy is given by:

$$W^B = \frac{\pi(e_p^B)^2 + (1-\pi)(e_u^B)^2}{2} + \pi v^B(\ell p) + (1-\pi)v^B(\ell u).$$

Aggregate welfare in the free market allocation is $W^0 = (\pi(1-\alpha)^2 + (1-\pi)(1-\alpha)^2\beta^2)/2 + \pi\alpha + (1-\pi)\alpha\beta$. At $\pi = 1$, $W^0 = (1-\alpha)^2/2 + \alpha > (1/2-\alpha)^2(1+2(1-\alpha))/(2(1-\alpha)^2) + \alpha = W^B$, where the inequality follows from $\alpha < \delta/2$. For $\pi = 0$, $W^0 = \beta^2(1-\alpha)^2/2 + \alpha\beta > \delta(1/2-\alpha)^2(\delta + 2\beta(1-\alpha))/(2(1-\alpha)^2) + \alpha\beta$, using that $\beta > \delta/2 > \alpha$.

The difference in welfare between a B policy and the free market is:

$$\begin{aligned} W^0 - W^B =& \frac{\pi((1-\alpha)^2 - (e_p^B)^2) + (1-\pi)((1-\alpha)^2\beta^2 - (e_u^B)^2)}{2} \\ & - \pi(v^B(\ell p) - \alpha) - (1-\pi)(v^B(\ell u) - \alpha\beta). \end{aligned}$$

That is, $W^0 > W^B$ if

$$\begin{aligned} &\pi((1-\alpha)^2 - (e_p^B)^2) + (1-\pi)((1-\alpha)^2\beta^2 - (e_u^B)^2) \\ &> \pi(1-\pi)2\alpha(2\delta - \beta - 1) + (1-2\alpha)\left(\pi(\pi e_p^B + (1-\pi)e_u^B\delta) + (1-\pi)(\pi e_p^B\delta + (1-\pi)e_u^B\beta)\right). \end{aligned}$$

Since $2\pi(1-\pi)(2\delta-\beta-1) < \pi(\pi e_p^B + (1-\pi)e_u^B\delta) + (1-\pi)(\pi e_p^B\delta + (1-\pi)e_u^B\beta)$ under our assumptions, $W^0 > W^B$ is implied by

$$\begin{aligned} &\pi((1-\alpha)^2 - (e_p^B)^2) + (1-\pi)((1-\alpha)^2\beta^2 - (e_u^B)^2) \\ &> (1-\alpha)\left(\pi(\pi e_p^B + (1-\pi)e_u^B\delta) + (1-\pi)(\pi e_p^B\delta + (1-\pi)e_u^B\beta)\right). \end{aligned}$$

Using the upper bounds on e_p^B and e_u^B from above a sufficient condition is:

$$\begin{aligned} (1-\alpha)^2 &\left(\pi(1 - \frac{1}{4}(\pi + (1-\pi)\delta)^2) + (1-\pi)(\beta^2 - \frac{1}{4}(\pi\delta + (1-\pi)\beta)^2)\right) \\ &> (1-\alpha)\frac{1-\alpha}{2}\left(\pi(\pi + (1-\pi)\delta)^2 + (1-\pi)(\pi\delta + (1-\pi)\beta)^2\right). \end{aligned}$$

Therefore:

$$\pi\left(1 - \frac{3}{4}(\pi + (1-\pi)\delta)^2\right) + (1-\pi)\left(\beta^2 - \frac{3}{4}(\pi\delta + (1-\pi)\beta)^2\right) > 0.$$

Rewriting the sufficient condition yields:

$$\pi + (1-\pi)\beta + 3\pi(1-\pi)(1-\delta)\left(1+\delta+\pi(1-\delta)\right)$$
$$-\,3(1-\pi)\pi(\delta-\beta)(\pi(\delta-\beta)+2\beta) > 0.$$

This becomes:

$$\beta + \pi(1-\beta) + 3\pi(1-\pi)\left(1-\delta^2-\pi(1-\beta)(2\delta-1-\beta)-2\beta(\delta-\beta)\right) > 0.$$

Since $\pi(1-\pi) < 1/4$, $2\delta - 1 - \beta < 1$, and $2(\delta-\beta) < 1$ the above condition must hold true under our assumptions and therefore $W^0 > W^B$. Moreover, since $e_p^B < e_p^0$ and $e_u^B < e_u^0$ $W^0 > W^B$ also implies that $Y^0 > Y^B$.

Second Best Policy

Given a policy $\rho(ab, ab')$ the payoffs of the different attributes are given by:

$$v(hp) = (2\rho(hp,hp) + \rho(hp,hu)\delta + \rho(hp,\ell p)/2 + \rho(hp,\ell u)\delta/2)/(\pi e_p),$$
$$v(\ell p) = (\rho(hp,\ell p)/2 + \rho(hu,\ell p)\delta/2 + 2\rho(\ell p,\ell p)\alpha + \rho(\ell p,\ell u)\alpha\delta)/(\pi(1-e_p)),$$
$$v(hu) = (2\rho(hu,hu)\beta + \rho(hp,hu)\delta + \rho(hu,\ell p)\delta/2 + \rho(hu,\ell u)\beta/2)/((1-\pi)e_u),$$
$$v(\ell u) = (\rho(\ell u,hp)\delta/2 + \rho(\ell u,hu)\beta/2 + \rho(\ell u,\ell p)\alpha\delta + 2\rho(\ell u,\ell u)\alpha\beta)/((1-\pi)(1-e_u)).$$

Since $\sum \rho(hp,.) = \pi e_p$ and similarly for the other attributes this leaves six choice variables.

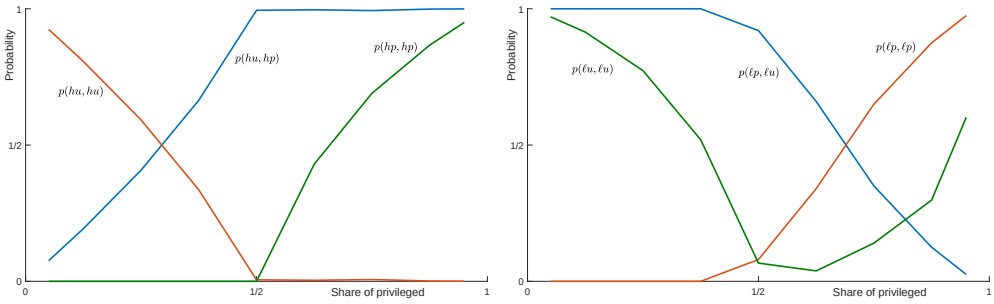

Figure 13: High (left) and low (right) achievers' matching probabilities in the second best.

We solved the problem numerically and Figure 13 shows the second best optimal matching probabilities for the parametrization used to generate all the figures ($\delta = .9$, $\beta = .6$, $\alpha = .2$). Comparing surplus values to those under an A policy and free market yields the numbers in the text.

B.3 Proofs for Section 5

Proof of Proposition 6

We first derive the competitive equilibrium. Payoffs for ℓu and hp who segregate are given by $v(\ell u) = \alpha\beta$ and $v(hp) = 1$. As stated above $-t(\ell p, hu) = t(hu, \ell p) \in [\beta - \delta/2, \delta/2 - \alpha]$ is determined by the relative scarcity of attributes hu and ℓp. Because of no arbitrage all colleges with the same support have the same transfers and composition so that we drop the subscript c. Agents' investments are given by $e_u^C = \delta/2 + t(\ell p, hu) - \alpha\beta$ and $e_p^C = 1 - \delta/2 + t(\ell p, hu)$.

Suppose $\pi(1 - e_p^C) < (1 - \pi)e_u^C$ first. Then $t(\ell p, hu) = \beta - \delta/2$ and:

$$e_u^C = (1 - \alpha)\beta \text{ and } e_p^C = 1 + \beta - \delta.$$

This regime occurs for $\pi < \frac{\beta - \alpha\beta}{\delta - \alpha\beta}$. $v(\ell p) = \delta - \beta$.

Second, suppose that $\pi(1 - e_p^C) = (1 - \pi)e_u^C$. This implies that $t(\ell p, hu) = (1 - \pi)\alpha\beta + (2\pi - 1)\delta/2$, and:

$$e_u^C = \pi(\delta - \alpha\beta) \text{ and } e_p^C = 1 - (1 - \pi)(\delta - \alpha\beta).$$

This may hold for $\frac{\beta - \alpha\beta}{\delta - \alpha\beta} \leq \pi \leq 1 - \frac{\alpha}{\delta - \alpha\beta}$. $v(\ell p) = (1 - \pi)(\delta - \alpha\beta)$.

Finally, if $\pi(1 - e_p^C) > (1 - \pi)e_u^C$, $t(\ell p, hu) = \delta/2 - \alpha$. Then

$$e_u^C = \delta - (1 + \beta)\alpha \text{ and } e_p^C = 1 - \alpha.$$

This regime occurs if $\pi > 1 - \frac{\alpha}{\delta - \alpha\beta}$. $v(\ell p) = \alpha$.

Note that $e_p^C / e_u^C \geq (1 - \alpha)/(\delta - (1 + \beta)\alpha)$, since both e_u^C and e_p^C increase in π at the same rate $\delta(1 - \alpha)$.

$\overline{A}$ Policy

Denote again by ρ the share of hu students who attend (hu, hp) univer-

sities under an $\overline{A}$ policy. Payoffs are given as:

$$
\begin{aligned}
v(hp) &= 1 - \frac{(1-\pi)e_u\rho}{(1-\pi)e_u\rho + \pi e_p}(1-\delta),\\
v(hu) &= \delta - \frac{(1-\pi)e_u\rho}{(1-\pi)e_u\rho + \pi e_p}(\delta-\beta),\\
v(\ell u) &= \alpha\beta,\\
v(\ell p) &= \alpha + \frac{(1-\pi)e_u(1-\rho)}{(1-\pi)e_u(1-\rho) + \pi(1-e_p)}(\delta/2 - \alpha - t(hu,\ell p)),
\end{aligned}
$$

where $t(hu,\ell p) > 0$ is the transfer that ℓp students pay in $(hu,\ell p)$ colleges. If hu students attend both (hu,hp) and $(hu,\ell p)$ colleges the transfer has to satisfy:

$$
\frac{\pi e_p}{(1-\pi)e_u\rho + \pi e_p}(\delta-\beta) = \frac{\pi(1-e_p)}{(1-\pi)e_u(1-\rho) + \pi(1-e_p)}(\delta/2 + t(hu,\ell p) - \beta).
$$

From the proof of Lemma 8 we know that colleges with both ℓp and hu form if and only if:

$$
\pi < \frac{(\delta-\beta-\alpha)e_u^A}{(\delta-\beta-\alpha)e_u^A + \alpha e_p^A} := \pi^*.
$$

Note that in this case, $t(hu,\ell p) = \delta/2 - \alpha$, i.e. ℓp will only obtain their segregation payoff. This is because supposing that $\rho < 1$ and $t(hu,\ell p) < \delta/2 - \alpha$ will lead to a contradiction of stability of the college equilibrium. Given ρ and $t(.)$ there are $\rho' > \rho$ and $t'(hu,\ell p) > t(hu,\ell p)$ such that both $v'(hu) > v(hu)$ and $v'(\ell p) > v(\ell p)$. To see this define $A = p_c'(hu,\ell p)/p_c(hu,\ell p) = (\pi(1-e_p) + (1-\pi)e_u(1-\rho))/(\pi(1-e_p) + (1-\pi)e_u(1-\rho'))$, and $B = t'(.)/t(.)$, and that the conditions become:

$$
\begin{aligned}
(AB-1)p_c(hu,\ell p)t(hu,\ell p) &> (A-1)p_c(hu,\ell p)(\beta - \delta/2)\\
(AB-1)p_c(hu,\ell p)t(hu,\ell p) &> (A-1)p_c(hu,\ell p)(\delta/2 - \alpha) + (B-1)t(hu,\ell p).
\end{aligned}
$$

Fixing B, implicitly defined by:

$$
(AB-1)p_c(hu,\ell p)t(hu,\ell p) = (A-1)p_c(hu,\ell p)(\beta - \delta/2) + \epsilon,
$$

the second condition becomes:

$$0 > (A-1)p_c(hu,\ell p)(\delta-\beta-\alpha) + (B-1)t(hu,\ell p) - \epsilon,$$

Using the definition of B:

$$0 > (A-1)p_c(hu,\ell p)(\delta-\beta-\alpha) + \frac{(A-1)(\beta-\delta/2-t(hu,\ell p)) + t(hu,\ell p) + \epsilon/p_c(hu,\ell p)}{Ap_c(hu,\ell p)} - \epsilon.$$

Since $t(hu,\ell p) < \delta/2-\alpha$ by assumption and $p_c(hu,\ell p) < 1$ there is $1 < A < 1/p_c(hu,\ell p)$ such the condition is satisfied. Hanece, for all $p_c(hu,\ell p) < 1$ and $t(hu,\ell p) < \delta/2-\alpha$ there are $p_c' > p_c$ and $t'(.) > t(.)$ such that a college that offers admission to both hu and ℓp with these parameters will attract a positive measure of both attributes, contradicting stability of $t(.)$ and $p_c(.)$. Hence, $t_c(hu,\ell p) = \delta/2-\alpha$ in a college market equilibrium with partial transferability.

Since $v(\ell p) = \alpha$ independently of π, and $v(\ell u) = \alpha\beta$, investments under an $\overline{A}$ policy will be strictly greater under partially transferability than under NTU for $\pi < \pi^*$. This is because $v(hu)$ and $v(hp)$ both strictly increase in ρ and for $\rho = 1$, both $v(hu)$ and $v(hp)$ are the same under partially transferable utiltiy and NTU. Hence, both privileged and underprivileged investments an $\overline{A}$ policy are higher when utility is partially transferable. Since both surplus and output strictly increase in both e_p and e_u, and $v(\ell p)$ and $v(\ell u)$ remain the same, the $\overline{A}$ policy will achieve higher surplus and output when utility is partially transferable than under NTU.

If $\pi \geq \pi^*$ our results from above carry over and optimal investments under an $\overline{A}$ policy are therefore

$$e_u = \frac{\pi e_p(\delta-\beta)}{\pi e_p + (1-\pi)e_u} + \beta(1-\alpha),$$

and

$$e_p = 1-\alpha-\frac{(1-\pi)e_u(1-\delta)}{\pi e_p + (1-\pi)e_u}.$$

Comparing this regime to the competitive equilibrium under partial transferability, note that payoffs for both u and p agents can be higher under the policy. Suppose $\pi = 1/2$, in which case $e_p^{\overline{A}} > e_u^{\overline{A}}$. Indeed $\pi^* < 1/2$ if $2\alpha > \delta-\beta$, and $e_p^C = 1+\beta-\delta$ and $e_u^C = (1-\alpha)\beta$ if $\beta(1-\alpha) > \delta-\beta$. Expected payoffs (and surplus) for u students are clearly higher under the

policy since $e_u^{\overline{A}} > e_u^C$ and $v^{\overline{A}}(\ell u) = v^c(\ell u)$. For p students expected payoff is higher under the policy if

$$(e_p^{\overline{A}})^2 + \alpha > (1 - (\delta - \beta))^2 + \delta - \beta.$$

Since $e_p^{\overline{A}} > (1 + \delta - 2\alpha)/2$, this must be true for δ sufficiently close to 1.

For the underprivileged: $v^{\overline{A}}(\ell u) = v^C(\ell u) = \alpha\beta$. Hence, payoff, surplus and investment and greater under the policy if $e_u^{\overline{A}} > e_u^C$. For $\pi \geq \pi^*$ side payments are not used and $e_u^{\overline{A}} = e_u^A$. For $\pi < \pi^*$, side payments are positive and $v^{\overline{A}}(hu) > v(hu)$, where $v(hu)$ denotes an hu's payoff under an $\overline{A}$ policy without side payments. Therefore $e_u^{\overline{A}} > v(hu) - \alpha\beta > e_u^C$. Therefore $e_u^{\overline{A}} > e_u^C$ in both cases and surplus, payoff and investment of the underprivileged are higher under the policy. This is obvious for $\pi < (\beta - \alpha\beta)/(\delta - \alpha\beta)$ since then $e_u^C = e_u^0 < e_u^A$. For higher π, $e_u^C \leq \delta - (1 + \beta)\alpha$. Since $e_u^{\overline{A}}$ increases in $e_p^{\overline{A}}$, for $e_p^{\overline{A}} > e_u^{\overline{A}}$:

$$e_u^{\overline{A}} > (1 - \alpha)\beta + \pi(\delta - \beta) > \delta - (1 + \beta)\alpha,$$

for $\pi > (\beta - \alpha\beta)/(\delta - \alpha\beta)$. For $e_u^{\overline{A}} > e_p^{\overline{A}}$ we have that

$$e_u^{\overline{A}} > (\delta - \beta)\frac{\delta - \beta - \alpha(1 - \beta)}{2\delta - \beta - 1} > \delta - (1 + \beta)\alpha.$$

Moreover, since $e_p^{\overline{A}} < 1 - \alpha$ (because $v^{\overline{A}}(hp) < 1$ and $v^{\overline{A}}(\ell p) \geq \alpha$) we have $e_p^{\overline{A}}/e_u^{\overline{A}} < e_p^C/e_u^C$.

References

Becker, G. S. (1973), 'A theory of marriage: Part i', *Journal of Political Economy* **81**(4), 813–846.

Bénabou, R. (1993), 'Workings of a city: Location, education, and production', *Quarterly Journal of Economics* **108**(3), 619–652.

Bénabou, R. (1996), 'Equity and efficiency in human capital investment: The local connection', *Review of Economic Studies* **63**, 237–264.

Bhaskar, V. and Hopkins, E. (2016), 'Marriage as a rat race: Noisy pre-

marital investments with assortative matching', *Journal of Political Economy* **124**(4), 992–1045.

Bidner, C. (2014), 'A spillover-based theory of credentialism', *Canadian Journal of Economics* **47**(4), 1387–1425.

Booth, A. and Coles, M. (2010), 'Education, matching, and the allocative value of romance', *Journal of the European Economic Association* **8**(4), 744–775.

Carrell, S., Sacerdote, B., Econometrica, J. W. and 2013 (2013), 'From natural variation to optimal policy? the importance of endogenous peer group formation', *Econometrica* **81**(3), 855882.

Chade, H., Eeckhout, J. and Smith, L. (2017), 'Sorting through search and matching models in economics', *Journal of Economic Literature* **55**(2), 493544.

Chiappori, P. (2017), *Matching with transfers: The economics of love and marriage*.

Cicalo, A. (2012), 'Nerds and barbarians: Race and class encounters through affirmative action in a brazilian university', *Journal of Latin American Studies* **44**, 235–260.

Clotfelter, C., Vigdor, J. and Ladd, H. (2006), 'Federal oversight, local control, and the spectre of "resegregation" in southern schools', *American Law and Economics Review* **8**(3), 347–389.

Coate, S. and Loury, G. C. (1993), 'Will affirmative-action policies eliminate negative stereotypes?', *American Economic Review* **5**, 1220–1240.

Cole, H. L., Mailath, G. J. and Postlewaite, A. (2001), 'Efficient noncontractible investments in large economies', *Journal of Economic Theory* **101**, 333–373.

de Bartolome, C. A. (1990), 'Equilibrium and inefficiency in a community model with peer group effects', *Journal of Political Economy* **98**(1), 110–133.

Dillon, E. W. and Smith, J. A. (2013), 'The determinants of mismatch between students and colleges', *NBER Working Paper Series* (Nr. 19286).

Durlauf, S. N. (1996*a*), 'Associational redistribution: A defense', *Politics & Society* **24**(2), 391–410.

Durlauf, S. N. (1996*b*), 'A theory of persistent income inequality', *Journal of Economic Growth* **1**(1), 75–93.

Epple, D. and Romano, R. E. (1998), 'Competition between private and public schools, vouchers, and peer-group effects', *American Economic Review* **88**(1), 33–62.

Faust, D. G. (2015), '2015 remarks at morning prayer', Office of the President, Harvard University. Retrieved from `http://www.harvard.edu/president/speech/2015/2015-remarks-morning-prayers`.

Felli, L. and Roberts, K. (2016), 'Does competition solve the hold-up problem?', *Economica* **83**(329), 172–200.

Fernández, R. and Galí, J. (1999), 'To each according to...? markets, tournaments, and the matching problem with borrowing constraints', *Review of Economic Studies* **66**(4), 799–824.

Fernández, R. and Rogerson, R. (2001), 'Sorting and long-run inequality', *Quarterly Journal of Economics* **116**(4), 1305–1341.

Fryer, R. G., Loury, G. C. and Yuret, T. (2008), 'An economic analysis of color-blind affirmative action', *Journal of Law, Economics, and Organization* **24**(2), 319–355.

Gall, T., Legros, P. and Newman, A. F. (2006), 'The timing of education', *Journal of the European Economic Association* **4**(2-3), 427–435.

Hong, L. and Page, S. E. (2001), 'Problem solving by heterogeneous agents', *Journal of Economic Theory* **97**, 123–163.

Hopkins, E. (2012), 'Job market signalling of relative position, or becker married to spence', *Journal of the European Economic Association* **10**(2), 290–322.

Hoppe, H., Moldovanu, B. and Sela, A. (2009), 'The theory of assortative matching based on costly signals', *Review of Economic Studies* **76**(1), 253–281.

Hoxby, C. M. and Avery, C. (2013), 'The missing "one-offs": The hidden supply of high-achieving, low-income students', *Brookings Papers on Economic Activity* (Spring 2013), 1–65.

Lang, K. and Lehman, J.-Y. K. (2011), 'Racial discrimination in the labor market: theory and empirics', *Journal of Economic Literature* **50**(4), 959–1006.

Legros, P. and Newman, A. F. (2007), 'Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities', *Econometrica* **75**(4), 1073–1102.

Lutz, B. (2011), 'The end of court-ordered desegregation', *American Economic Journal: Economic Policy* **3**(2), 130–168.

Nöldeke, G. and Samuelson, L. (2015), 'Investment and competitive matching', *Econometrica* **83**(3), 835–896.

Orfield, G. and Eaton, S. E. (1996), *Dismantling desegregation: the quiet reversal of Brown v. Board of education*, The New Press, New York, NY.

Peters, M. and Siow, A. (2002), 'Competing pre-marital investments', *Journal of Political Economy* **110**, 592–608.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for dartmouth roommates', *Quarterly Journal of Economics* **116**(2), 681–704.

Weinstein, J. (2011), 'The impact of university racial compositions on neighborhood racial compositions: Evidence from university redistricting', *mimeo* .