College Diversity and Investment Incentives*

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

November 2017

Abstract

We study diversity policies, such as affirmative action in college admission or inclusion policies within the college walls, in the presence of local peer effects. If students are constrained in making side payments within peer groups, the free market allocation displays excessive segregation relative to the first-best, generating excessively disparate pre-college investments. Effective diversity policy must overcome market forces within as well as across college boundaries and combine admission and inclusion policies. Policies that engender diversity affect pre-college investment incentives. 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 student cross-subsidization by colleges.

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) 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 been a source of both controversy and challenge.

The controversy comes from the costs and benefits of the obvious policy response to the problem of implementing a desired level of diversity at the university: just impose it! If for instance one would like to replicate the diversity in the population, a system of quotas could be imposed (when it is public funded). For private universities, if peer effects due to diversity increase the average reputation of the university, its officials may provide affirmative action policies to achieve this goal, e.g., by putting aside a number of positions for students from a given background. This solution has to face however the usual criticism that policies that force diversity may distort the incentives to invest in education prior to entering the university, either for those students who are favored by the policy or by the others. We have here the two main elements that fuel policy debates: diversity may be desirable from socio or political objectives (equity, diversity or righting past wrongs) but it comes at an economic cost.[1] In this paper we take issue with this tradeoff and argue that because of non-transferabilities affirmative action in universities is often beneficial from an economic perspective.

The challenge is to find ways to implement such affirmative action policies. In academic and policy debates, even when the desirability of intervention in the market outcome is taken for granted, there are open questions about what is the most effective way to accomplish it. College administrators who have nonetheless been enthusiastic supporters of the diversity policies in their universities express dissatisfaction with the levels actually achieved. Others note that segregation *within* the college gates may be undermining the very ends college diversity is meant to achieve. Harvard president Drew Faust expresses 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).

[1]It also comes at a legal cost. There have been numerous legal challenges to university affirmative action and other diversity policies, along with vigorous defenses by the universities. See for instance the *amicus curiae* briefs in the Supreme Court cases Fisher vs. Texas 2012 and 2015, jointly submitted by many of the best US universities. We discuss the academic literature below.

This last concern underscores the evident importance of peer groups for the college experience. They matter not only for what one learns while there, but also for what one earns afterward. Effective peer groups are often small — much smaller than the university one attends. Thus while the admission policy of a university may go some way toward accomplishing a given diversity objective, market forces may continue to exercise their influence within its boundaries, enabling effective segregation to persist, and preventing effective *inclusion*. A university's "local" policy regarding free association — in effect its ability and willingness to enforce diversity at the level of peer groups — may ultimately be the crucial determinant of diversity within its boundaries. This suggests that a well thought-out diversity policy must address two separate issues: bring within college walls a diverse population of students and, second, induce diverse peer groups. As we show the two policies are jointly needed. Focusing on only diversity at the admission level may not induce diverse peer groups, but focusing only on diversity at the peer group level will induce entry by colleges that favor less diversity at the admission level.

This paper offers a theoretical examination of the effects of diversity policies on the investments prior to college, on the distributions and levels of income and surplus. We employ a "NTU-networking-with-investment" model, which allows us to focus on two salient features of the college marketplace. The first is the already-noted relatively small size of peer groups within a college. The second feature of the college marketplace is that the benefits students accrue from attending college and interacting with peers cannot easily be transferred among them via a price system. There are many reasons for this non-transferability (NTU) including, but not limited to, moral hazard, social norms, regulations, or limited financial ability to make transfers based on lifetime benefits. In such a non-transferable world, policies of tax-subsidies may be ineffective or imperfect instruments for achieving the desired goals set by a planner or the college officials.

These two features make the resulting free-market allocation of students into peer groups within colleges potentially problematic. The model illustrates that achieving diversity at the college level may require intervention at the "local" (peer group) level, something that is potentially more challenging than simply altering "gatekeeping" (admissions) policies. A promising approach could be to exploit mechanisms that are already in place in many universities, but not regularly used for promoting peer group inclusion: assignment to dorms and dorm rooms, to peer mentoring groups, and to tutorial classes.[2]

Our analysis builds on 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.[3] At the time they are admitted to college, agents have attributes that reflect their *background* (privileged or underprivileged) and their early education *achievement* (high or low). Privilege and high achievement increase both one's own and one's peers' payoffs to attending college. While background is exogenous, achievement is the result of an earlier investment. We assume that local peer effects are strongest when peers have diverse backgrounds. 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 networking policies.

Colleges control admission and can therefore determine the characteristics of students within their gates. We contrast two post admission policies. In the *laissez-faire policy* students are free to choose their local networks. In the *random inclusion policy*, students cannot choose their local network and are instead linking with other students in proportion to their representation in the college.

Under NTU, the laissez-faire policy within the college is characterized by full segregation in achievement and background within college walls, independently of the admission policy. This implies that the equilibrium choice of investment by individuals is independent of college compositions, or on policy interventions at the admission level.[4] For instance, colleges segregated by

[2]Empirical studies often achieve identification of peer effects — which appear to be non-negligible — by the fact that universities randomize assignments to dorm rooms (Sacerdote, 2001; Stinebrickner and Stinebrickner, 2006; Kremer and Lavy, 2008) or classes (Lerner and Malmendier, 2013); the point is such assignments, or simple modifications thereof, could also be used as policy tools.

[3]It would be straightforward to extend 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. See our discussion in the conclusion.

[4]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

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 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).

Under a random networking policy, segregation is avoided with the college if a diverse group of student is admitted. But universities could escape the inclusion policy by being selective in the characteristics of students admitted. Hence inclusion policies by themselves cannot change equilibrium outcome. A combination of constraints on admission – inducing a diverse enough body of students – and inclusion – inducing heterogeneous local networks is necessary for a change in the way individuals link in colleges and generate peer effects.

Now, such policies may distort the incentives to invest in education prior to entering college, both for those students who are favored by the policy and perhaps more importantly, those who are not. And the fear may be that the negative effects on investments will undermine the positive effects of having more diverse networks. As we show, this trade-off may be misconstrued. Indeed, since the free market generates the "wrong" match, investment incentives are also distorted. Though rematch policies cannot directly address the market imperfections, they may provide an instrument for correcting distortions in both the match and investments; properly designed, they can raise aggregate output and investment, reduce inequality, and increase welfare. Diversity policies may be beneficial both for equity and efficiency.

Our inclusion policies have been defined above. For admission policies, we mimic actual practice, and consider two types of policies, and look at their effects when inclusion policies are in place. We first consider "achievement blind" policies that only focus on replicating the diversity of backgrounds in the population, a typical example being "busing" (to be sure, in the U.S. at least, this sort of policy has been largely confined to primary and secondary

therefore of different types of matches) in the economy.

schools rather than higher education). While this type of policy may generate higher aggregate surplus than free market, it guarantees low achievers a "good" match, and high achievers a "bad" one, with sufficient probability as to significantly depress investment incentives.

We then consider an "affirmative action" policy, which is defined as one that conditions the priority for admission given to an underprivileged on achievement: among the underprivileged, only the high achievers are considered candidates for admission. 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. Numerical simulations indicate that our affirmative action policy can come very close to the optimal re-matching policy.

The qualitative results persist if we assume some limited transferability where privileged agents have sufficient wealth to make transfers into the college marketplace, but underprivileged have limited ability to pay. Naturally, the free market outcome changes; instead of global segregation, privileged *low* achievers match with underprivileged *high* achievers. This still fails to be welfare maximizing, and affirmative action policies help improve aggregate performance. Indeed this case underscores the difference between gatekeeping and local policies: a university that admits only high achievers will attract only the privileged, the underprivileged will go elsewhere since they benefit from being in peer groups with low ability privileged students.

The paper proceeds as follows. Following this paragraph we review the literature on matching and policy making in the face of excessive segregation. In Section 2 we lay out the model framework. In Section 3 we show that segregation obtains when agents have no wealth and that it leads to distorted investment incentives with respect to the ideal situation where agents have large initial wealth. This opens the door for diversity policies to be surplus and welfare enhancing and we show that this is the case in Section 4. In fact, when the benefits from diversity are high in terms of total surplus and welfare, an affirmative action policy – a priority to high ability underprivileged for admission – coupled with an inclusion policy is close to the second-best policy. We allow in Section 5 some transferability among students but limited since the underprivileged are wealth constrained and have difficulties borrowing; we comfort the benefits of using affirmative action policies in this case. We conclude in Section 6. All proofs and calculations not in the text can be found in the appendix.

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 and also adds to the theoretical literature on matching with endogenous types.

Literature on widespread externalities and rematch policies. 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 small, 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"). In their model, affirmative action is consistent with two types of equilibria. In the "bad" affirmative action equilibrium, although employment of the underprivileged may increase, beliefs do not change, leaving investment incentives and wages unchanged or reduced. But in the "good" equilibrium, as in our (unique) equilibrium, affirmative action provides an incentive for the underprivileged to invest because they believe they will actually get a job; meanwhile employers observe that they are productive, so beliefs are consistent.

One would expect after such a policy had been in place for a while that the benefits would persist if it were subsequently removed. This seems inconsistent with empirical observations for colleges: suspending affirmative action policies that have been in place for a while have often triggered reversion to the pre-policy status quo.[5] Since evolving beliefs are not part of our NTU framework, our model easily explains this empirical regularity.

Existing work tends to evaluate the performance of policies with respect to the objective of colleges, for instance, as in Fryer et al. (2008) who 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 evaluate the general equilibrium effects of these policies, e.g., rarely discuss the effects 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, which are among the questions we analyze in this paper.

Carrell et al. (2013) designed an experiment in which half the freshmen at the United States Air Force Academy were assigned to peer groups including *low ability* students and find that students formed sub-groups on the basis of ability within such peer groups, negating the intended objective of inducing more interaction between high and low ability students. In our model such segregation between low and high ability students is in fact efficient (this would not be the case for segregation of low ability privileged and high ability underprivileged however). Their result nevertheless illustrates the challenges in inducing heterogeneity in local peer groups. The segregation at local peer groups has been documented also in high schools .

[5] 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.

Literature on matching. The theoretical literature on matching 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.[6] 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. 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 recently 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.[7]

Booth and Coles (2010) compare different marriage institutions in terms of their impact on matching and investments. Gall et al. (2006) analyze the impact of timing of investment on allocative efficiency. Several studies consider investments before matching under asymmetric information (see e.g., Bidner, 2014; Hopkins, 2012; Hoppe et al., 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.

[6] 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).

[7] Nöldeke and Samuelson (2015) provide a general analysis of matching with non transferability and investment prior to the match.

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 which we normalize: $\sum_{ab \in A} q(ab) = 1$. Student s may also have a wealth endowment ω_s. In the NTU case, ω_s is "small" for all agents, implying that transfers are insufficient to change the matching outcome obtained when $\omega_s = 0$. We will also consider the idealized first best case where ω_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. Achieving h with probability e requires an investment in education of e at individual cost $e^2/2$. In the market agents are fully characterized by their *attributes* ab and their wealth.

Colleges

Students choose to attend one of n colleges. A college c has size k_c. An admission policy of a college is a distribution q_c over A; when $q_c(ab) = 0$, it is known that the college does not admit students of attribute ab, but when $q_c(ab) > 0$, the college will admit this proportion of students. The choice of admission policy by the college may be affected by affirmative action 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 early social connections will generate (like referrals for jobs). The pattern of social interactions 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 may reflect constraints put in place 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 the college could force students to match randomly in dorms, in which case $p_c(ab, a'b') = q_c(ab)q_c(a'b')$. The way students match in peer groups must be consistent with the distribution of characteristics of admitted students.

Definition 1. p_c is *consistent* given q_c whenever 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 policies q_c to be easier to implement than inclusion policies p_c. For instance, limiting admission to students with particular attributes (hence using $q_c(ab) = 0$ for some attribute ab) is easy to arrange, but 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 would lead 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 the affirmative action (quota) policy, sat at the front (see also Carrell et al., 2013.) Nevertheless, colleges can put in place policies (for roommates' allocation, for tutoring, for class attendance) that will increase the probability that students of different characteristics match. We call such policies "inclusion policies".

Finally, within colleges, a monetary transfer policy $t_c : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$ may be put in place, where $t_c(ab, a'b')$ is the payment made by a student with attribute ab when matched to a student with attribute $a'b'$; if negative, $t_c(a'b', ab)$ is the payment received by $a'b'$ when matches with ab. Such transfers could be centralized by the college; for instance, if the college anticipates that students will match following p_c, it is equivalent for the college to collect a payment of $\sum_{a'b'} t_c(ab, a'b')p_c(ab, a'b')$ from students of characteristic ab and to give an amount $r_c(a'b', ab)$ to students of characteristic $a'b'$ when they meet with a student of characteristic ab (for instance if $a'b'$ serves as a tutor for a student ab).

These transfers are subject to limited liability:

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

A college c can now be defined by the 4-tuple (k_c, q_c, p_c, t_c): (k_c, q_c) describes the set of students who are admitted (how many of them and their relative shares of attributes) while (p_c, t_c) describes the within college matches and transfers.

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

Outputs

The output generated by students with attributes ab and $a'b'$ who are matched is given by:

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

(For simplicity, we ignore the possibility that students may be unmatched; even if a student could have a positive payoff from being unmatched, assuming that $y(ab, ab)$ is larger than twice the payoff that a student of attribute ab would have if unmatched insures that in equilibrium no positive measure of students is unmatched.)

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 obviously 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, $g(u,p) - g(u,u) > g(p,p) - g(p,u)$) whenever $\delta - \beta > 1 - \delta$, 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 inclusion 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. There are many reasons to suspect that diversity in backgrounds is indeed 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 student who is privileged 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). Appendix B discusses alternate assumptions on the output functions.[8]

[8]Throughout we assume that students perceive the payoff function correctly. It is conceivable that in reality they underestimate the value of diversity; for instance we could suppose that the "true" payoff $\hat{y}(\cdot,\cdot)$ satisfies $\hat{y}(ap, a'u) > \hat{y}(ap, a'p)$, while for the perceived payoff $y(ap, a'u) < y(ap, a'p)$ as specified above. An alternate interpretation is that some sort of dynamic inconsistency, like hyperbolic discounting, leads them to behave as if they have the preferences we specify. In either case, the market outcome and positive effects

2.1 Timing

The timing in the model economy is as follows.

(1) Policies for admission and inclusion, if any, are 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, possibly constrained by admission policy.

(5) Within the college agents choose which peer(s) to interact with, possibly constrained by the inclusion policy.

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

2.2 Equilibrium

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

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

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

In equilibrium ,an agent of attribute ab chooses a college from the admissible set C^* to maximize her expected utility *conditional* on being admitted. Letting $I_c(ab)$ be the 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^*, the agent 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 (k_c, q_c, p_c, t_c) such that the following conditions hold:

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

of policy are unchanged, but the case for policy intervention becomes arguably even more compelling.

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

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

A consequence of stability is that colleges make zero profit. If a college makes a positive profit, then there exists $ab, a'b'$ such that $t_c(ab, a'b') + t_c(a'b', ab) < 0$, that is what one student pays is inferior to what the other receives. But then a new college could enter, with the same values for k_c, q_c, p_c and offer a new transfer scheme that will be preferred by all types in the college, for instance if R is the profit, give a lump sum a transfer of to each admitted student (hence the new transfer is $t_c(ab, a'b') - q_c(ab)R$).

Existence of a q-equilibrium is immediate from previous results on the core of infinite economies with finite coalitions Kaneko and Wooders (1986, 1996), but we will provide a constructive proof.

Investment. The anticipation of payoffs $v(ab|C^*)$ for all possible q-equilibria will influence the investments made by agents before their entry in the college admission market. Our assumption that attributes in the peer group match 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 agents $i \in [0, \pi)$ have background p and all agents in $i \in (\pi, 1]$ have background u. If the aggregate investment level of agents 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. Hence, agents of the same background b choose the same education investment e_b. We can therefore 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 exists a college market

$q(\mathbf{e})$-equilibrium C^* such that for any $b = u, p$.

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

Definition 4. An *equilibrium* is an investment $\mathbf{e}$ together with a college market $q(\mathbf{e})$-equilibrium such that $\mathbf{e}$ is an equilibrium investment given C^*.

There are in general multiple equilibria. Trivially, if college (k_c, q_c, p_c, t_c) is part of an equilibrium, then two colleges $(k_c/2, q_c, p_c, t_c)$ could be formed instead and also be part of an equilibrium. We will see other reasons for multiplicity below. Nevertheless, equilibria *outcomes* are essentially unique in the sense that each attribute obtains the same expected payoff in all possible equilibria.

3 Free Market with Non-Transferabilities and Investment Distortions

Before discussing the positive and normative effects of re-matching policies, it is useful to contrast the matching pattern and the investment levels obtained in the free market situation where agents have no wealth (or "little" wealth as we will see) with an ideal situation in which agents have no financial constraints and a price system exists for transferring utility at the peer group level. We consider this idealized situation below.

3.1 Free Market with Non-Transferabilities

In such an environment where transfers are not possible, a student in a pair $(ab, a'b')$ obtains payoff $y = f(a, a')g(b, b')$; the Pareto frontier for a match $(ab, a'b')$ consists therefore of a single point and $t_c(ab, a'b') = 0$. Our assumptions imply that the payoffs to each student in a match are given by the following matrix. The free market equilibrium 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 has admits hp students; within this college hp cannot obtain more than 1 in any match and will segregate; since $\beta > \delta/2$, hu, if admitted, will also segregate since they cannot attract hp in a match; now, because $\delta < 1$, ℓp, if admitted, will also segregate. This precludes having in equilibrium a positive probability

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 $(ab, a'b')$

of $(ab, a'b')$ peer groups, 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.[9]

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, at the college level or within colleges through peer group interaction. One could argue that because it may be difficult for students *not to encounter* students of different attributes when there is a diverse student body, that the likely free market outcome is for colleges to be segregated by attributes.

Equilibrium payoffs are uniquely determined, however:

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

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 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.

[9]Note that segregation will be the case whenever the underprivileged have wealth $\omega_u < 1 - \delta$ and the privileged have wealth $\omega_p < \beta - \delta/2$. Then still hp students strictly prefer being matched to hp with probability one to any other match obtaining the maximum transfer, and hu students strictly prefer being matched to hu with probability one to hu than any convex combination of hu and ℓp with the maximum transfer.

3.2 Free-Market with Full Transferability (First-Best)

Utility is fully transferable between partners in a match $(ab, a'b')$ when they can share the total output

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

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

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

The maximum transfer an individual is willing to make is equal to $y(ab, a'b')$, which corresponds to his life time earnings, which for most people is a degree of magnitude larger than the fees requested for attending the college. 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.[10] Because of this "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 peer group choice 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 shares $q(ab)$. The structure of payoffs and the stability conditions lead to the following observations.

(i) $p_c(hp, \ell u) = 0$: $(hp, \ell u)$ matches cannot occur in a first best allocation. Indeed, in an $(hp, \ell u)$ peer group hp agents lose more compared to their segregation payoff than ℓu students gain: the average surplus in matches (hp, hp) is 1, and $\alpha\beta$ in $(\ell u, \ell u)$ matches. The average surplus in a match $(hp, \ell u)$ is $\delta/2 < 1/2$, which is less than what hp students obtain in segregation.

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

[10] 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.

(iii) If agents of a given achievement match together, surplus is higher if backgrounds are diverse. Indeed, note that condition (DD), is equivalent to $2z(ap, au) > z(ap, ap) + z(au, au)$, implying that segregation in background is surplus inefficient.

(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 matching $(hu, \ell p)$ and segregating hp, than matching (hp, hu) and segregating ℓp: in the former case, total surplus is $2\delta+2$, compared to $4\delta+2\alpha$ in the latter case. Hence, in any equilibrium, all $(hu, \ell p)$ matches will be exhausted and matches (hu, hp) will form only if there is an excess supply of hu agents.

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 in the following to the set of α satisfying the following condition:

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

Lemma 1. *Under* (4)*, a first best allocation exhausts all possible* (hp, hu) *matches, then all* $(hu, \ell p)$ *matches, and then all* $(\ell p, \ell u)$ *matches, while all other remaining attributes segregate.*

Under full transferability, all equilibria must be efficient. Indeed, if

Figure 1 shows the possible efficient matching patterns under full transferability depending the desirability of diversity. The plain arc indicates the first priority matching, the dashed arc indicates the second priority potential match, once the first priority matches are exhausted, and the ellipsis matches when these second matches are exhausted. Again, the first best allocation specifies social interactions, but leaves open the precise 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, the matches $\ell p, hu$ will exhaust ℓp students and students of type ℓu could segregate in a different college than the other types since $p_c(\ell u, ab) = 0$ for all $ab \neq \ell u$.)

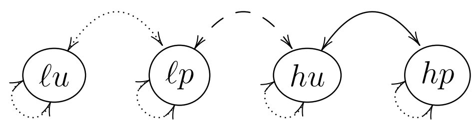

Figure 1: Ex-ante peer networks with TU

Again the equilibrium definition allows different outcomes in terms of colleges' compositions. However some configurations must arise. For instance, all *hp* students cannot be segregated since they should interact with *hu* students within a college, hence a top college must admit both *hu* and *hp* students. Depending on the distribution q, students of type ℓp should also be in the same college as *hu* students. But it is also possible that all colleges reflect the population measures of attributes and $p_c(.)$ and $t_c(.)$ satisfy the equilibrium conditions.As above investments depend on the market premium for high achievement $v^*(hb) - v^*(\ell b)$. The payoffs for attributes $v^*(ab)$ will depend on relative scarcity, which in turn will depend on the initial measure of privileged π and achievable surplus $z(ab, a'b')$.

The following statement summarizes the properties of TU equilibrium investment levels when there is a high diversity benefit.

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^* *being U shaped and* e_u^* *being an inverted U shape.*

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 agents, peer group returns reflect scarcity: scarce agents in the matching market can claim a high share of the total peer group return. For this reason the scarcity of privileged as measured by π will affect the returns from college and therefore the incentives to invest in education. By contrast, when there is no possibility of transfer the peer group returns will not reflect scarcity: there will be segregation and therefore the return of an attribute is independent of the distribution of attributes, hence of π. This explains why privileged agents 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.

A more precise characterization of the investment outcomes is offered in the following proposition that is illustrated in Figure 2.[11]

Proposition 1. *The underprivileged never over-invest and under-invest if* $\pi > \frac{(1-\alpha)\beta}{1+(1-\alpha)\beta}$. *The privileged over-invest for* $\frac{2\alpha+(1-\alpha)(2\delta-1)}{2\alpha+(1-\alpha)2\delta} < \pi < \frac{1-(1-\alpha)(2\delta-1)}{2-(1-\alpha)(2\delta)}$, *in which case there is both over-investment at the top and under-investment at the bottom of the background distribution.*

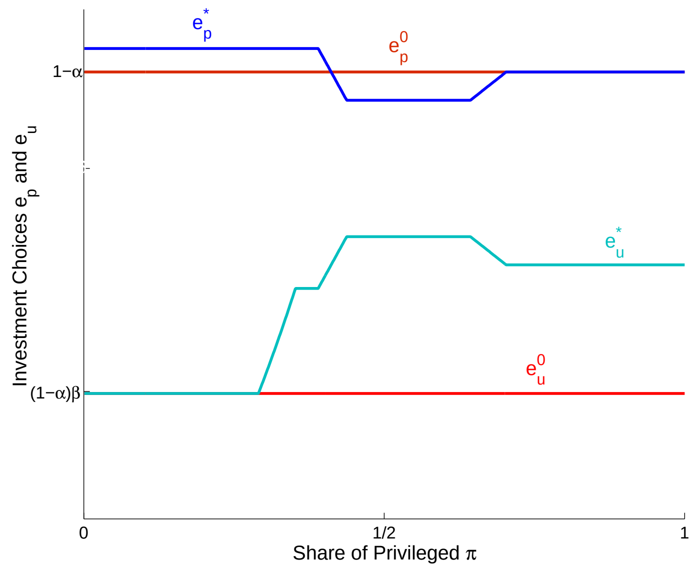

Figure 2: Education investments: NTU (e^0) vs TU (e^*)

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

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

inefficiency, but also generates investment distortions, a dynamic inefficiency. Specifically, there will tend to be insufficient investment by the under-privileged; as for the privileged, their investment will be insufficient or excessive depending on whether they are a small enough minority. As we shall see, this suggests that the possible discouragement effects on the privileged that diversity policies introduce may actually be desirable.

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. An *admission policy* will affect the shares of each attribute admitted to a college c, $q_c(ab)$, for instance requiring them to be the same as in the population. An *inclusion policy* will affect the matching probabilities $p_c(ab, a'b')$ within the college. While admission rules that aim to generate a certain student composition on campus are relatively easy to implement (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 in our environment any policy that only affects admission or within college allocation will not be effective.

Proposition 2. *Under NTU, using exclusively either an admission policy or an inclusion policy replicates the free market.*

As we have seen, absent the possibility of transfers, there is full segregation within the college, that is, independently of the distribution q_c, $p_c(ab, ab) = 1$ for admitted students ab. A possible equilibrium is for colleges to be segregated at the admission level, that is colleges are such that they admit only one type of student ($q_c(ab) = 1$). An affirmative action policy would force colleges to admit a more diverse body *if students of different background apply for admission*, which seems an improvement over segregated colleges. However absent an inclusion policy, we will still have $p_c(ab, ab) = 1$ within such colleges, hence the expected payoffs of agents is the same, and the investments will be the same. In fact, hu students may as well go to a segregate college rather than going to a college hp in which they will gain admission; for them the end outcome will be the same! As an illustration of the force towards segregation within colleges, Cicalo 2012 describes how students of different backgrounds spatially segregated in the classroom under a quota policy in admission at a big university in Rio de Janeiro. Another example is provided by the self-selection of students into dormitories, resulting in segregated student houses or segregated corridors, while an inclusion policy would allocate students on the basis of a lottery system.

Absent an affirmative action policy for admission, an inclusion policy within colleges will trigger sorting across colleges, rendering inclusion within colleges quite pointless. Indeed, if the inclusion policy is such that $p_c(hp, ab) > 0$ for some type $ab \neq hp$, a segregated college admitting only hp students will offer a higher payoff to hp types. Replicating this argument, the effect of inclusion policies will be to make segregated colleges more attractive, and a way to bypass the inclusion policy.

Therefore, to be effective, policies must constraint both the way admissions are made *and* the way students match within colleges. But on which basis these policies should be designed? Consider simple "background blind" policies based on achievement only, applied to both stages: admission is based on achievement only ($q_c(au) > 0 \Leftrightarrow q_c(ap) > 0$) and matching in the college are on the basis of achievement ($p_c(\ell b, hb') = 0$), as in honors programs. Colleges are not allowed to use admission rules or organize interactions within college based on backgrounds. Because hp students strictly prefer to be with other hp students, $p_c(hp, hp) = 1$ is an equilibrium match that does not violate the inclusion policy and there will be the same matching outcome as without the background blind policies. Hence this simple policy is also ineffective.

Proposition 3. *Background blind policies replicate the free market.*

For the remainder of this section we focus on two other policies that affect both admission and within college interactions, and that lead to equilibria that differ from the free-market equilibrium. We model inclusion policies on campus in the simplest conceivable way by imposing random matching within college. An earlier version of the paper considered the case that colleges have full control over social interactions on campus; the results are qualitatively the same but quantitative effects are stronger since random matching is not in general the "optimal" inclusion policy.

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

The adoption of such inclusion policies 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.

For college admission we focus on two extreme policies. First, we will consider an "achievement-blind" policy, which assigns students to colleges by background without regard to achievements, in order to replicate the population frequency: the probability that a u matches with a p is just π. This policy effectively pins down admission quotas.

Second, we study "affirmative action," which gives priority to the underprivileged over privileged students in admission, but only for students with high achievement level. By granting priority rights this policy (unlike quota policies) affects college composition indirectly by changing the market outcome.

Because a large part of the efficiency of the match is linked to the achievement element of the attributes, an achievement-blind policy tends often to perform worse than an affirmative action policy. Studying these polar cases allows some inference on intermediate ones, like scoring policies where a score reflecting both achievement and background determines priority.

At the risk of repetition, in our discussion below, we assume that colleges use an inclusion policy, that is that there is random matching among a college's students.

4.1 Achievement-Blind Policy for Admission

Several real-world policies are essentially achievement blind. Post 1968, public European universities often did not condition admission on achievement beyond the basic requirement of finishing high school; formally, this is akin to an assignment rule that randomly integrates peer groups in background, ignoring achievement. In the U.S., this type of policy has been mainly restricted to primary and secondary education. Possibly the most prominent example is the use of "busing" to achieve high school inclusion, which operated mainly by redesigning school districts to reflect aggregate population measures. Other examples are the inclusion of school catchment areas in Brighton and Hove, U.K.; reservation in India to improve representation of scheduled castes and tribes; the Employment Equality Act in South Africa, under which some industries such as construction and financial services used employment or representation quotas; or the SAMEN law in the Netherlands (until 2003).

Definition 6. Under an *achievement Blind policy* (denoted B policy) a college cannot discriminate on the basis of achievement for admission and has to admit u students if it admits p students. Within colleges there is random matching.

Note that this policy precludes, in equilibrium, the formation of colleges admitting only h students. h students would value such colleges since the inclusion policy will limit the matches to hu, hp students while it would lead to matches with $\ell u, \ell p$ students if ℓ students are admitted. However, such a segregated college cannot refuse to admit ℓ students, and since these students benefit from matching with h students, segregated colleges on the basis of achievement cannot be part of an equilibrium under the B policy.

Hence, a B policy limits colleges to offering compositions $q_c(ab)$ equal to population quotas. A B policy is thus best understood as a quota policy that departs from the free market outcome of full segregation and randomly reassigns agents to match the expected share of privileged students at each college to their population measure π.

The following statement describes the resulting assignment of students.

Lemma 4. *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 3 describes the possible matches.

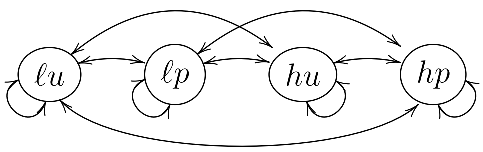

Figure 3: Ex-ante peer networks under a B policy.

Because this policy allows both (hu, hp) matches and $(\ell p, hu)$ matches, it may be beneficial for increasing surplus *if investment in achievement is not important*, for instance when the distribution of types is exogenous. However, because the assignment rule does not depend on achievement, investment incentives are likely to be depressed compared to the free market.[12] This may explain why these policies have been mainly used at the primary or secondary levels rather than at the university level where prior investment in human capital is arguably more important.

The following statement uses Lemma 4 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.*

That is, a B policy is indeed subject to the classic equity-efficiency tradeoff that seems to guide much of the policy discussion. Reducing outcome

[12]Both ℓ and h agents of background b have the same chance of being matched to an h agent of background b'

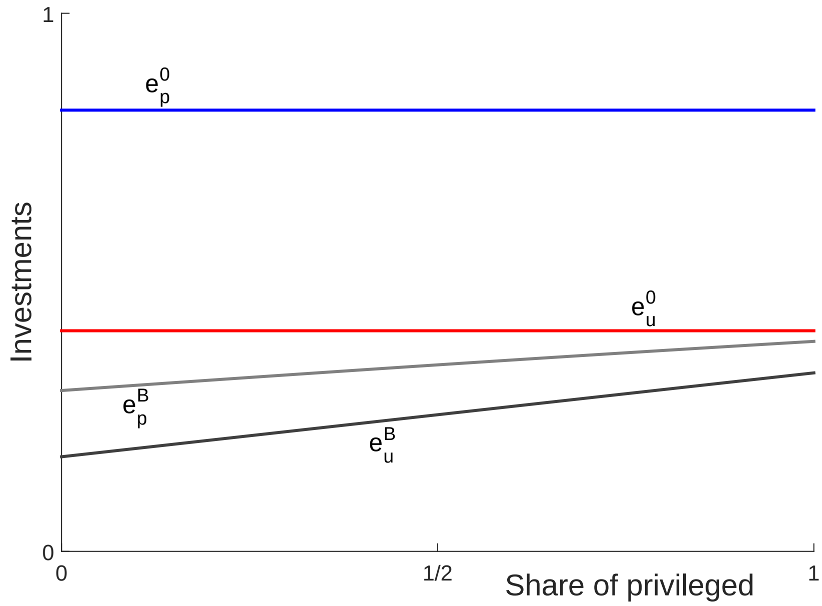

Figure 4: Education investments: B policy (e^B) and laissez-faire (e^0).

inequality in the economy comes at the cost of undesirable incentive effects depressing levels of investment and output: both the privileged and the underprivileged are discouraged relative to the free market outcome because higher investment does not increase the probability of obtaining a better match, see Figure 4. In fact, when the privileged are a minority, a B policy can reverse the background gap in investment so that $e_p^B < e_u^B$. This comparative statics exercise assumes that when π varies, both δ, β stay constant, which may be a strong assumption in general.

4.2 Affirmative Action Policy

We examine now the case where precedence is given for an underprivileged candidate over a privileged competitor for students of *high* achievement level only.[13] Formally, affirmative action is a priority for the underprivileged for positions at a given level of achievement in segregated universities. It is widely used (for instance, the reservation of places for highly qualified minority students at some *grandes écoles* in France, like Sciences Po Paris, the "positive equality bill" and *Gleichstellung* in the public sectors in the U.K.

[13]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: one could award priority for the underprivileged also among the low achievers, priority could be stochastic, perhaps closer to actual scoring rules.

and Germany).

Definition 7. Under an *affirmative action policy* (denoted A policy) any underprivileged student with achievement h is guaranteed a place at any college that also admits privileged students with achievement h. Within colleges there is random matching.

Under this policy no arbitrage has to ensure that students with attribute hu strictly prefer their equilibrium college to any other college with $q(hu) > 0$ or $q(hp) > 0$. 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 two lemmas provide two key properties of the college market equilibrium.

The support of types in a college c is the set of types ab for which $q_c(ab) > 0$.

Lemma 5 (Symmetric Equilibria)**.** *Under an inclusion policy, for almost all parameter values, if the equilibrium supports of types for two colleges coincide, they have the same distribution of types.*

Proof. 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. □

Lemma 6. *Under an A policy in equilibrium all ℓu and ℓp students will segregate into colleges, i.e., $p_c(\ell u, \ell u) = 1$ and $p_c(\ell p, \ell p) = 1$. All hu and hp will match into colleges with $q_c(hp) = \pi e_p/(\pi e_p + (1-\pi)e_u)$ and $q_c(hu) = 1 - q(hp)$.*

Proof. Notice that under random matching within the college equilibria with free entry are symmetric: all colleges with the same support of attributes have also the same student composition.

Moreover, under full NTU all students prefer to be matched with hp, then hu, then ℓp and then ℓu students. Since ℓp and ℓu students have no priority they will segregate as in the free market equilibrium. 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. □

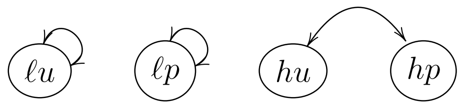

Figure 5: Ex-ante peer networks under an A policy.

The ex-ante equilibrium networks in Figure 5 is consistent with the following college configuration. There are three types of colleges: colleges who admit only ℓu students, those admitting only ℓp students and finally those admitting hp, hu students; in the last type of colleges, hu and hp match randomly since there is an inclusion policy in place. As under the B policy optimal individual investment levels will depend on the match an agent expects to obtain, and thus on relative scarcities. Since the privileged only have to accept underprivileged matches if they have high achievement level, privileged investments will be less depressed than under the B policy. The following proposition states this and other properties of aggregate outcomes under an A policy; details are in the appendix.

Proposition 5. *Under an A policy the underprivileged invest more than under free market ($e_u^A > e_u^0 > e_u^B$), and the privileged less ($e_p^0 > e_p^A > e_p^B$). Inequality of investments between backgrounds is smaller under the A policy than under free market, and the education gap between backgrounds may reverse if $\delta > 1 - \alpha(1-\beta)$ and the measure of privileged is close to 1. Aggregate investment is higher; aggregate output and welfare are higher if diversity is desirable enough.*

Not only does an A policy crowd out privileged investment by less than a B policy, but also underprivileged investment is boosted compared to the free market, as illustrated in Figure 6. 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 ℓu college, and the agent's loss is greater than under a B policy since he would have a chance there to be matched with h students. Therefore the expected returns to investment are now conditional on integrating in backgrounds if successful.

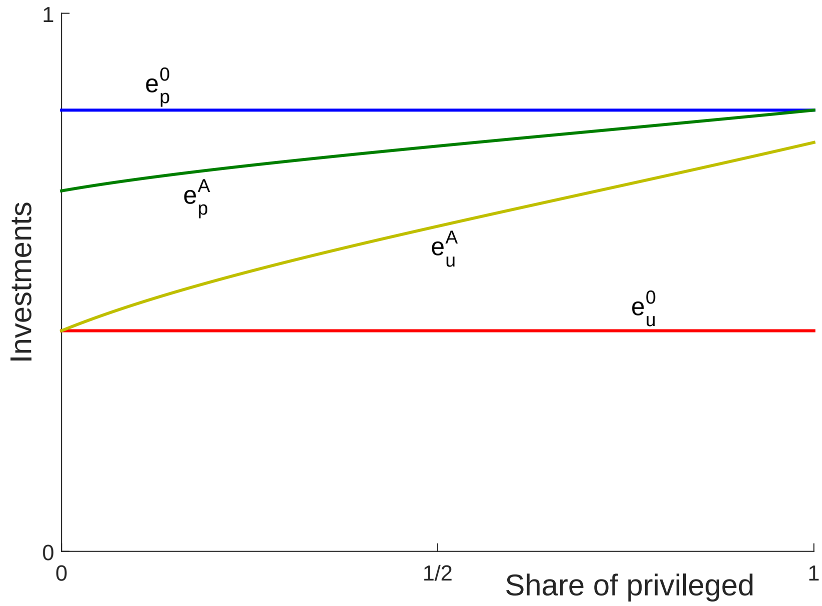

Figure 6: Education investments using an A policy.

This encourages the underprivileged and discourages the privileged, and, if diversity is desirable – that is condition (DD) holds – the aggregate effect on investment is positive. If diversity is desirable or backgrounds are distributed unevenly also aggregate output is higher.

4.3 Aggregate Effects

The two policies of re-match considered above differ substantially in terms of their position in the trade-off between static and dynamic concerns, that is between achieving more efficient sorting ex post (when attributes have been realized) and maintaining investment incentives by rewarding investments adequately through the match. Policies that emphasize replicating population frequencies of backgrounds in each peer group (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 that have 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 firms for the underprivileged. Figure 7 illustrates the differences in aggregate performance.[14]

[14]Even if the proportions of attributes is given, that is even if one is not concerned about investment incentives, an affirmative action policy dominates an achievement blind policy, and also free market: the A policy foregoes $(hu, \ell p)$ matches but avoids many

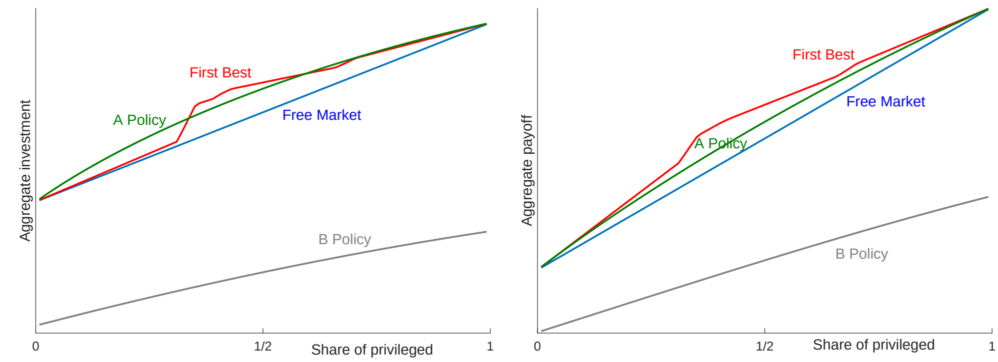

Figure 7: Aggregate investments (left) and aggregate payoff (right).

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

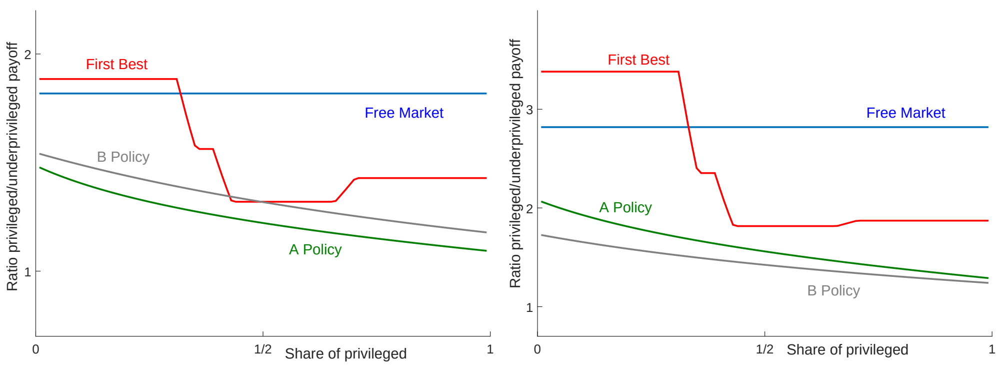

Figure 8: Inequality of investments (left) and payoffs (right).

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.

---

other surplus decreasing matches, like $(hp, \ell u)$ that arise under a B policy. Obviously, if incentives are ignored, the "naive" policy that replicates the first best match distribution under TU performs even better than the A policy.

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.

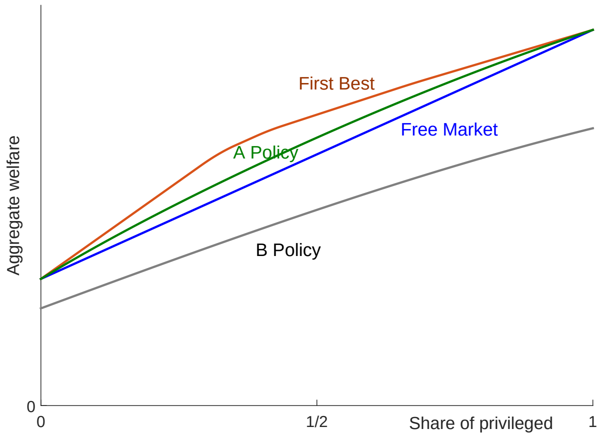

Figure 9: Total Surplus

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 $\delta = 0.9$).

Proposition 6 (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 A policy induces strictly higher total surplus than the free market with NTU if $\delta > \hat{\delta}(\pi)$.

4.4 Second-Best Surplus Maximizing Policy

While figure 9 suggests that the A policy is in fact close to the surplus maximizing policy for high values of δ, it may be of independent interest to compute the second-best optimal policy, that is when a planner has full control over the way agents will match, hence controls the matching probabilities $p(ab, a'b')$ subject to feasibility. The optimization problem of a planner 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. 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 second best policy closely resembles an A policy for our parameters. See Appendix A for details. In fact the A policy realizes more than 97% of the gains in surplus that the second best policy achieves (for $\delta = .9$, $\beta = .6$, and $\alpha = .2$, used for all figures).

The set of policies also includes scoring polices that give priority to students based on scores: convex combinations of achievements and backgrounds. For instance, one could give "grade subsidies" based on ethnicity (as the university of Michigan until 2003) or on whether a student attended a public high school (used in college admission in Brazil), or comes from a disadvantaged neighborhood.

Another policy is one that would replicate the first best matching, that is as in Figure 1. This "naive" policy faces a similar trade-off as the B policy: while maximizing the static gains from re-matching ex post, it falls short of optimizing the incentives, because the *payoffs*, which are still constrained by NTU, cannot replicate the TU outcome.[15] Instead it may be better for the planner to approximate the TU investment incentives by generating convex combinations of NTU payoffs that differ from those that would be accomplished by the naive policy – the second best policy takes full advantage of

[15] Calling this policy "naive" is a bit of misnomer, as it has a serious practical drawback: it would require considerable sophistication on the part of the policy maker to compute the (counterfactual) TU outcome!

this possibility. Indeed, an A policy can sometimes outperform the naive policy, and the naive policy is the A policy when $\pi > 1/2$ (details in the Appendix).

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 ω_b, depending on their background b. Plausibly, privileged background is associated with higher wealth. As mentioned in footnote 9, for $\omega_u < \alpha(1-\delta)$ and $\omega_p < \beta - \delta/2$ 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

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

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

The next statement follows directly from this observation.

Lemma 7. *Under* (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 equilibrium matching pattern. The under-

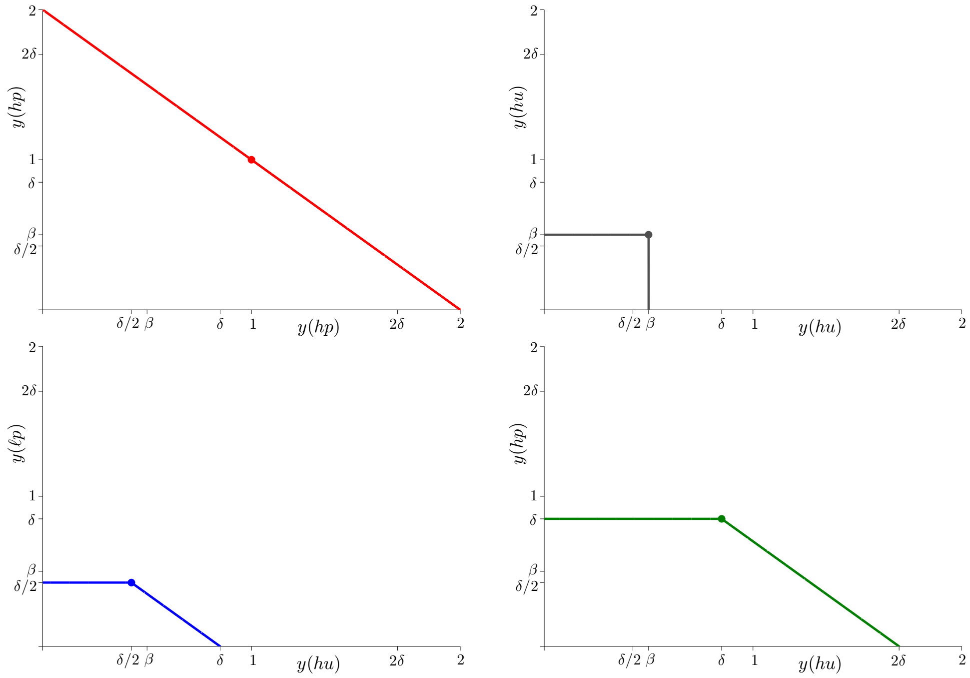

Figure 10: Possible distribution of payoffs in (hp, hp) and (hu, hu) peer groups (top) and $(\ell p, hu)$ and (hp, hu) peer groups (bottom) when individuals can make lump-sum transfers but the underprivileged face borrowing constraints.

privileged match with the privileged, but only in $(hu, \ell p)$, not in (hu, hp) peer groups, and the elite (hp, hp) peer groups are solely populated by the privileged, which seems to resonate well with the evidence.[16] Observe that a background blind policy at the admission level will replicate also the free market outcome, as in Proposition 2. Indeed, colleges can have two types of admission policies: admit only high achievers, or admit both. hp will self select into the first type of colleges while $hu, \ell p$ will self-select into the second type of colleges, leaving the ℓu to segregate, replicating the free market outcome.

[16]For instance, Dillon and Smith (2013) find evidence for substantial mismatch in the U.S. higher education system, in the sense 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 of students is lower than their own achievement and seem less costly, in marked 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.

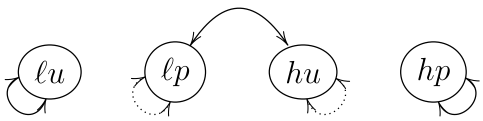

Figure 11: Equilibrium colleges q_c and matches p_c with transfers

Note that with (5), an A policy (in conjunction with random matching within the colleges) yields matches between hu and hp students, since hu students have priority in college admission. However, ℓp student would still pay to interact with hu students and thus hu can choose between better peers and better money. If hu are scarce and obtain a hp match with near certainty, ℓp students prefer segregating to compensating. That is, for high shares of the privileged the A policy with partial transferability coincides with the one under NTU.

Lemma 8. *Under* (5)*, a college market equilibrium under an A policy yields colleges with both hu and hp students, and colleges with both hu and ℓp students if the population share of hp students is low enough. ℓu students attend segregated colleges.*

As in the case without side payments, an A policy encourages investment by the underprivileged, since underprivileged high achievers are rewarded with access to privileged high achievers. By contrast, when side payments are possible an A policy may encourage investments by students of *both* backgrounds. This is because limited wealth limits competition among ℓp's, thereby giving rents to privileged low achievers. An A policy depresses these rents 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 who are forced to match with the underprivileged, so that investment incentives for the privileged increase. Indeed for intermediate π this encouragement effect is so strong that the expected payoff ex post of a privileged student is higher under an A policy, if diversity is desirable enough (δ sufficiently large).

Proposition 7. *Suppose Conditions (DD) and* (5) *hold. An A policy induces higher investment and payoffs for the underprivileged, and reduces the investment gap between backgrounds. If δ is high enough, an A policy induces higher investment for each background, and for intermediate π also higher payoffs for both backgrounds.*

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

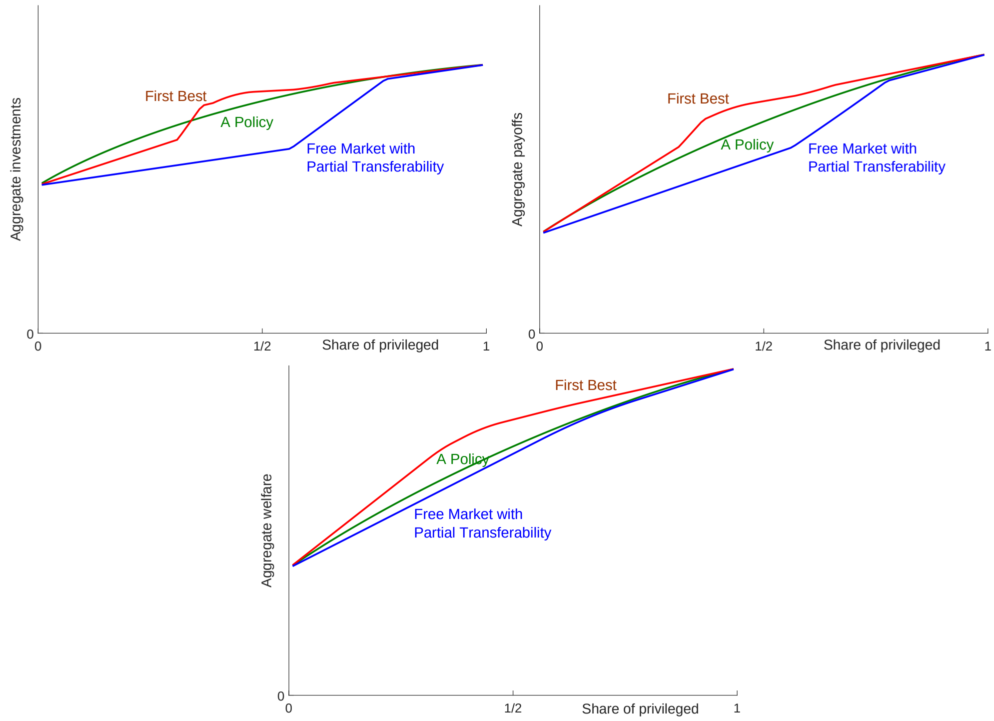

Figure 12: Aggregate investments (left), income (right), and surplus (bottom) when $\omega_p > \delta/2 - \alpha, \omega_u = 0$

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

However, because hu students have the right but are not compelled to match 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 peer groups (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.[17]

[17]In practice, such transfers could be effectuated through donations by the hp's (or their parents) to the scholarship funds of the second tier peer groups; c.f. the recent controversy over donations by the Koch brothers 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/).

For instance, two hp and ℓp students each, who are matched into (hp, hu) and $(\ell p, \ell u)$ groups, could jointly offer side payments to the two hu students to achieve a rematch into groups $\{(hp, hp), (hu, \ell p), (hu, \ell p), (\ell u, \ell u)\}$. 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 obtain an hp match, and ℓp students would pay at most $\delta/2 - \alpha\delta$ to replace their ℓu with an hu match. That is, given an 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 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 *matching*, 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 A policy only led to rematch, while the p's have the same investment incentives whether or not the rematch is effected – thus the 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 matches.

A second 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) matches by giving the hu's sufficient wealth to make the side payment needed to attract an hp (in practice this is a voucher or scholarship, since the wealth given to the hu cannot be spent arbitrarily, and in practice might take the form of reduced or waived tuition along with a living stipend). 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.

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 terms of 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.

6 Conclusion

An 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 peer group are a local public good, and students have limited means with which to make transfers, we show that the free market will indeed generate excessive segregation, and as a consequence, under-investment by the underprivileged and over investment by the privileged. These outcomes happen even if students know that they benefit from diversity and even if the benefit is at the level of their (small) peer group: NTU at the local level is the source of all distortions.

We study two simple policies that integrate backgrounds at the peer group level. One matches population measures without considering achievement and one gives priority to one background only conditional on achievement. The latter policy typically improves on the free market outcome (or on a policy that conditions only on achievement, which replicates the free market) in terms of aggregate investment, output, surplus, and inequality.

Though not exhaustive, the set of policies we examine covers the two extremes in terms of conditioning inclusion on achievement, allowing us to uncover considerable differences in the consequences for investment incentives, suggesting that conditioning on achievement is desirable. Moreover, numerical simulations show that this policy may in fact come close to a second best. While of interest, the question of the "optimal policy" in general settings is best left to future research. This quest will require us to compute complex contingencies, which will raise the issue of its practical implementation. Our focus on policies that are actually used by policymakers yields a convincing economic rationale for the use of such policies, when students' ability to make side payments is constrained and diversity is desirable.

We have introduced the possibility of transfers and as long as underprivileged have limited wealth or difficulties borrowing, vouchers or grants have limited success in generating diversity. Vouchers are feasible but not a market equilibrium. Similarly, need-blind policies are feasible but diversity would require that smart underprivileged hu apply for admission while the smart privileged hp are also willing to apply: as we argue, this would require that hp actually pay less than hu for otherwise they would segregate.

One of the novelties of our approach is to focus on the composition of peer groups within colleges as the source of excessive segregation. Making this as 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 will have no effect. By contrast if part of the college premium is due to quality of faculties and facilities, or to widespread externalities, admission policies can play an important role. 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 even if there is segregation within peer groups. A full analysis of the interaction between admission policies and local diversity policies is an interesting question 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 than for the privileged. 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 match with the underprivileged, will reduce their investment less. A pertinent observation is that investments often happen in environments such as primary and secondary 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 matching policies in another, as well as the complementarity or substitutability of rematch policies on sequential markets.

Finally, we have focused on how students match into colleges, where rigidities arise naturally from local public goods and borrowing constraints. Our results extend to other settings as well, e.g., the labor market. Contractual arrangement among the members of a firm are often designed to address agency problems. This typically results in a second best contract, inducing substantial non-transferabilities among firm members. This can be sufficient to generate excessive segregation and opens the door to a similar analysis of the aggregate effects of affirmative action policies in the labor market. Firms, unlike universities, have little reluctance to exercise managerial authority in the assignment of employees into teams or work groups. Thus, whatever the relative interest of firms and universities in achieving diversity, firms arguably have a more powerful array of instruments to get there.

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.

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.
- $\pi \geq \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_p^* = 1-\alpha$.

Similarly, for

- $\pi \leq \frac{(1-\alpha)\beta}{1+(1-\alpha)\beta}$, $e_u = (1-\alpha)\beta$.
- $\frac{(1-\alpha)\beta}{1+(1-\alpha)\beta} < \pi < \frac{(1-2\alpha)\delta}{1-(1-2\alpha)\delta}$ 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.
- $\pi \geq \frac{2-2\delta(1-\alpha)-\alpha}{2-2\delta(1-\alpha)}$, $e_u = (1-\alpha)(2\delta-1)$.

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.

A.2 Proofs for Section 4

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 $e_p^B \leq (1/2-\alpha)/(1-\alpha) < \delta/2 < (1-\alpha) = e_p^0$, and $e_p^B < \delta/2 < \delta - \alpha \leq e_p^A$ using that $1-\delta < \alpha < \delta/2$. For the underprivileged $e_u^B \leq \delta(1/2-\alpha)/(1-\alpha) < \beta(1-\alpha) = e_u^0 < e_u^A$. 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

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

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:

$$\begin{aligned}&\pi+(1-\pi)\beta+3\pi(1-\pi)(1-\delta)\left(1+\delta+\pi(1-\delta)\right)\\&-3(1-\pi)\pi(\delta-\beta)(\pi(\delta-\beta)+2\beta)>0.\end{aligned}$$

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 $Y^0>Y^B$.

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:

$$Y \xrightarrow{r2-2r4} \begin{bmatrix} 1 & \delta & 1/2 & \delta/2 \\ 0 & 0 & \delta(1/2-2\alpha) & \beta(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix}$$

$$\xrightarrow{r1-2r3} \begin{bmatrix} 0 & 0 & 1/2-2\alpha & \delta(1/2-2\alpha) \\ 0 & 0 & \delta(1/2-2\alpha) & \beta(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix}$$

$$\xrightarrow{r2-\delta r1} \begin{bmatrix} 0 & 0 & 1/2-2\alpha & \delta(1/2-2\alpha) \\ 0 & 0 & 0 & (\beta-\delta)(1/2-2\alpha) \\ 1/2 & \delta/2 & \alpha & \alpha\delta \\ \delta/2 & \beta/2 & \alpha\delta & \alpha\beta \end{bmatrix}$$

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

Solving $x'\hat{Y} = 0$, the first equation is $x_3/2 = 0$ which implies implies $x_3 = 0$. As long as $\beta \neq \delta^2$, $\alpha \neq 1/4$ and $\beta \neq \delta$, ub 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

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

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

and

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

Rewriting and dividing the second by the first equation yields:

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

This implies that e_p increases in e_u at a rate of less than unity. Using this fact and the above expression on (A.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:

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

Lower investment inequality (i.e., $e_p^A/e_u^A < 1/\beta = e_p^0/e_u^0$) follows directly from the expressions for e_p^A and e_u^A above. Notice that $e_u^A > e_p^A$ 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 $v^0(\ell b) = v^B(\ell b)$ for $b = u, p$, and both $e_p^A < e_p^0$ and $e_u^A > e_u^0$.

For the remaining assertions start with aggregate investment. It is higher under the A policy than under laissez faire if

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

Using the expressions above this becomes

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

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

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

Using (A.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 $v^A(\ell u) = v^0(\ell u)$ and $v^A(\ell p) = v^0(\ell p)$ the welfare comparison reduces to:

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

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

$$(1-\pi)(e_u^A - e_u^0)(e_u^A + e_u^0) > \pi(e_p^0 - e_p^A)(e_p^0 + e_p^A). \quad \text{(A.5)}$$

Using the expressions for e_u^A and e_p^A from above:

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

Using this expression on (A.5) yields

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

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

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

Since $\frac{e_p^A}{e_u^A}$ decreases in π, this ratio is bounded above by $(\delta-\alpha)/(\beta(1-\alpha))$, a sufficient condition for $W^A > W^0$ for all $\pi \in (0,1)$ is:

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

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 A policy.

Proof of Proposition 6

In the proof of Proposition 4 we showed that $W^B < W^0$ for all π. The second part of the proposition is proved as part of the proof of Proposition 5.

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.

We solved the problem numerically and Figure 13 shows the second best optimal matching for the parametrization used to generate all the figures ($\delta = .9$, $\beta = .6$, $\alpha = .2$). The broken lines correspond to matching probabilities under an A policy for comparison. That is, an A policy is indeed very close to

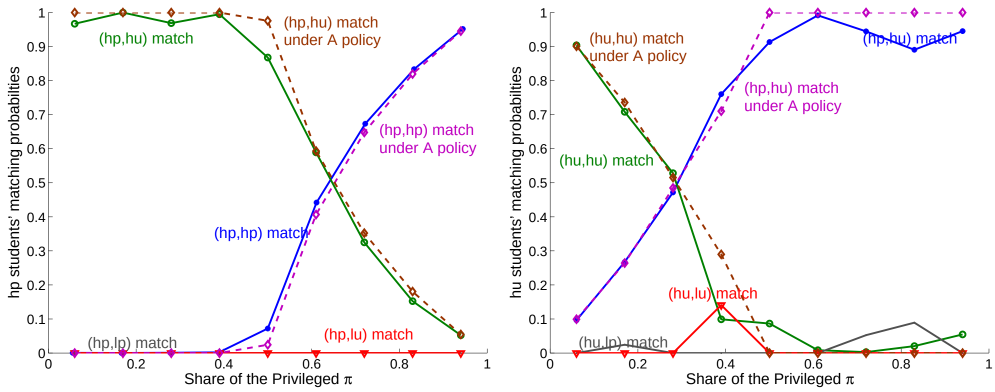

Figure 13: hp (left) and hu (right) students' matching probabilities in the second best.

second best for this particular parametrization when $\pi \geq 1/2$.[18] Comparing surplus values to those under an A policy and free market yields the numbers in the text.

A.3 Proofs for Section 5

Proof of Lemma 7

In an within college equilibrium with $q_c(ab) > 0$ for all attributes hp and ℓu students segregate with tuition fees $t(ab, ab) = 0$. 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 $q_c(hp) = q_c(\ell p)$.

In a college market equilibrium no arbitrage 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, $p_c(hp, \ell p) = p_{c'}(hp, \ell p)$ implies $t_c(hp, \ell p) = t_{c'}(hp, \ell p)$. 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 $p_c(hu, \ell p) = \frac{\pi(1-e_p)}{(1-\pi)e_u + \pi(1-e_p)}$.

[18] This result becomes more pronounced when δ is closer to 1. For low δ the second best may take the form of a naive policy, details are available from the authors.

Proof of Lemma 8

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

Proof of Proposition 7

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 $\delta(1-\alpha)$.

A **Policy**

Under an A policy denote by ρ the share of hu students who attend (hu, hp) universities. 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:

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

To have colleges with both hu and ℓp there must exist some share $\rho \in (0,1)$ such that $v(\ell p) > \alpha$, that is $t(hu,\ell p) < \delta/2-\alpha$. Using this, for colleges with both ℓp and hu to form, there must be $\rho \in (0,1)$ such that:

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

Since the left hand decreases in ρ colleges with both ℓp and hu form if and only if:

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

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

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

and

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

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

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

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

For the underprivileged: $v^{AC}(\ell u) = v^C(\ell u) = \alpha\beta$. Hence, payoff, surplus and investment and greater under the policy if $e_u^{AC} > e_u^C$. For $\pi \geq \pi^*$ side payments are not used and $e_u^{AC} = e_u^A$. For $\pi < \pi^*$, side payments are positive and $v^{AC}(hu) > v^A(hu)$ so that $e_u^{AC} > e_u^A$. Therefore $e_u^A > e_u^C$ 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^A increases in e_p^A, for $e_p^A > e_u^A$:

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

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

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

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

B Appendix: Generalized Surplus Function

Denote attributes by $s \in \{\ell u; \ell p; hu; hp\}$, endowed with a natural order, satisfying $\ell u < \ell p, hu$ and $hp > hu, \ell p$. Let $z(s, s')$ be monotone in its arguments ($z(s, s') > z(s, s'')$ if $s' > s''$).[19] Assume that $z(hp, hp) < 2$ to permit easy interpretation of investments as probabilities. The functional form $z(s, s') = 2f(a, a')g(b, b')$ satisfies these assumptions.

Diversity is desirable, that is, for $s = ab$ and $s' = a'b'$ with $b \neq b'$

$$2z(s, s') > z(s, s) + z(s', s'). \tag{DD}$$

This corresponds to the case of $2\delta > 1 + \beta$ in the functional form used above. Note that this property does not restrict the surplus function with respect to the composition of achievements ℓ and h, in particular decreasing and increasing differences are possible.

$z(.)$ satisfies *complementarity* of diversity and returns to education if

$$2[z(hu, s) - z(\ell u, s)] \geq z(hu, hu) - z(\ell u, \ell u) \text{ for } s \in \{hp, \ell p\}. \tag{C}$$

For this general surplus function, our OTUB result generalizes when (DD) and (C) hold.

Proposition 8. *Suppose properties (DD) and (C) hold.*
(i) There is $\underline{\pi} > 1/2$ *such that for all* $\underline{\pi} < \pi \leq 1$ *under free market privileged agents over-invest* ($e_p^* > e_p^T$)*, and underprivileged agents under-invest* ($e_u^* < e_u^T$)*.*
(ii) If $\pi < \underline{\pi}$ *and* $z(hu, hu) - z(\ell u, \ell u) < 1$ *there is under-investment by the underprivileged* ($e_u^* \leq e_u^T$)*. Under-investment is strict if additionally* $z(hu, hu) - z(\ell u, \ell u) < 2(z(hu, \ell u) - z(\ell u, \ell u))$*.*

The threshold $\underline{\pi}$ is given by $\underline{\pi} = \frac{1}{2(z(hp,hp) - z(hp,\ell p))}$ if $2z(hp, \ell p) > z(hp, hp) + z(\ell p, \ell p)$ and by $\underline{\pi} = \frac{1}{z(hp,hp) + z(\ell u,\ell u) - 2z(\ell p,\ell u)}$ otherwise.

Proof. Because of property (DD) under TU there cannot be positive measures of both matches $(ab, a'b)$ and $(ab', a'b')$. Hence, for any composition of achievements (a, a') the TU allocation exhausts all possible matches with background composition (u, p).

[19] A weaker form of monotonicity, $z(s, s') < \max\{z(s, s); z(s', s')\} \leq z(hp, hp)$ for all $s \neq s'$ is sufficient.

(i) Start by examining the case of $\pi e_p^T > 1/2$, i.e., oversupply of hp agents under TU. In this case $v(hp) = z(hp, hp)/2$ and $v(hu) = z(hp, hu) - z(hp, hp)/2$ by property (DD).

Suppose $(hp, \ell p)$ matches occur in equilibrium then $v(\ell p) = z(hp, \ell p) - z(hp, hp)/2$ and $e_p^T = z(hp, hp) - z(hp, \ell p)$ yielding the condition

$$\pi > 1/2(z(hp, hp) - z(hp, \ell p)).$$

Moreover, $e_p^T = z(hp, hp) - z(hp, \ell p) > (z(hp, hp) - z(\ell p, \ell p))/2 = e_p^*$ since $(hp, \ell p)$ matches occur (and thus are preferred by both hp and ℓp agents to segregation). $v(\ell u) = z(hp, \ell u) - v(hp)$ by property (DD), since $(hp, \ell p)$ matches occur. This means $e_u^T = z(hu, hp) - z(\ell u, hp) > (z(hu, hu) - z(\ell u, \ell u))/2 = e_u^*$ by property C.

Suppose $(hp, \ell p)$ matches do not occur in equilibrium. Then $(\ell p, \ell u)$ matches occur in equilibrium by property (DD). If $\pi(1-e_p) < (1-\pi)(1-e_u)$ then $v(\ell u) = z(\ell u, \ell u)/2$ and $v(\ell p) = z(\ell p, \ell u) - z(\ell u, \ell u)/2 > z(\ell p, \ell p)/2$. Hence, $e_p^T = z(hp, hp)/2 + z(\ell u, \ell u)/2 - z(\ell p, \ell u) < e_p^*$. $e_u^T = v(hu) - z(\ell u, \ell u)/2 > (z(hu, hu) - z(\ell u, \ell u))/2 = e_u^*$. Using these expressions reveals that $\pi e_p^T > 1/2$ implies $\pi(1 - e_p) < (1 - \pi)(1 - e_u)$. Therefore oversupply of hp agents and absence of $(hp, \ell p)$ matches is only consistent with $\pi(1 - e_p) < (1 - \pi)(1 - e_u)$.

(ii) If there are $(\ell u, \ell u)$ matches $v(\ell u) = v(\ell u, \ell u)/2$. If $z(hu, hu) + z(\ell u, \ell u) < 2z(hu, \ell u)$ there cannot be (hu, hu) matches as well. Therefore $w(hu) > z(hu, hu)/2$ and $e_u^T > e_u^*$. Otherwise ℓu agents' payoffs are determined by the equilibrium matches $(\ell u, s)$ yielding $v(\ell u) = z(\ell u, s) - v(s)$ for some skill level $s \in \{hu; \ell p; hp\}$. $w(hu) \geq z(hu, s) - v(s)$ with strict inequality if matches (hu, s) do not occur in equilibrium. Suppose there is $s \in \{hp; \ell p\}$ so that $(\ell u, s)$ matches occur in equilibrium, then by Property (C) $e_p^T = z(hu, s) - z(\ell u, s) > [z(hu, hu) - z(\ell u, \ell u)]/2 = e_u^*$. Otherwise all ℓu agents must be matched to hu, which requires $e_u^T > 1/2$. If $z(hu, hu) - z(\ell u, \ell u) < 1$ this implies $e_u^T > e_u^*$. □

A Policy vs. Free Market

The following proposition provides an analogue to Proposition 6, stating that surplus under an A policy is higher than under free market if δ is close enough to 1.

Proposition 9. *Aggregate surplus under an A policy is higher than under free market if $z(hp, hu)$ is sufficiently close to $z(hp, hp)$.*

Proof. As shown above there is full segregation in an equilibrium under free market with investments:

$$e_p^0 = \frac{z(hp, hp) - z(\ell p, \ell p)}{2} \text{ and } e_u^0 = \frac{z(hu, hu) - z(\ell u, \ell u)}{2}.$$

Total surplus under free market is

$$\begin{aligned} S^0 = &\pi \frac{(z(hp, hp) - z(\ell p, \ell p))^2}{8} + \pi \frac{z(\ell p, \ell p)}{2} \\ &+ (1 - \pi) \frac{(z(hu, hu) - z(\ell u, \ell u))^2}{8} + (1 - \pi) \frac{z(\ell u, \ell u)}{2}. \end{aligned}$$

Under an A policy both ℓp and ℓu agents segregate, so that $v^A(\ell p) = z(\ell p, \ell p)/2$ and $v^A(\ell u) = z(\ell u, \ell u)/2$. This means total surplus is higher under the A policy if

$$\begin{aligned} \pi \frac{(e_p^A)^2}{2} + (1 - \pi) \frac{(e_u^A)^2}{2} > &\pi \frac{(z(hp, hp) - z(\ell p, \ell p))^2}{8} \\ &+ (1 - \pi) \frac{(z(hu, hu) - z(\ell u, \ell u))^2}{8}. \end{aligned}$$

Since h types' wages depend on relative scarcity of background two different cases may arise. The first is that $(1 - \pi)e_u > \pi e_p$. Then $v^A(hp) = z(hp, hu)/2$ and

$$v^A(hu) = \frac{\pi}{(1 - \pi)e_u} \frac{z(hp, hu) - z(\ell p, \ell p)}{2} \frac{z(hp, hu) - z(hu, hu)}{2}.$$

This implies that

$$\begin{aligned} e_u^A = &\frac{z(hu, hu) - z(\ell u, \ell u)}{4} \\ &+ \frac{1}{2} \sqrt{\frac{(z(hu, hu) - z(\ell u, \ell u))^2}{4} + \frac{\pi}{1 - \pi}(z(hp, hu) - z(\ell p, \ell p))(z(hp, hu) - z(hu, hu))}. \end{aligned}$$

Using this the condition $(1 - \pi)e_u > \pi e_p$ becomes

$$\pi \le \frac{1}{2} \frac{z(hp, hu) - z(\ell u, \ell u)}{z(hp, hu) - [z(\ell u, \ell u) + z(\ell p, \ell p)]/2}.$$

Comparing surplus, $S^0 < S^A$ if

$$\begin{aligned}&\left(\frac{z(hp,hp)-z(\ell p,\ell p)}{z(hp,hu)-z(\ell p,\ell p)}\right)^2\\&< 1+\frac{z(hp,hu)-z(hu,hu)}{z(hp,hu)-z(\ell p,\ell p)}+\frac{1-\pi}{\pi}\left(\frac{z(hu,hu)-z(\ell u,\ell u)}{(z(hp,hu)-z(\ell p,\ell p))}\right)^2\\&\qquad\times\sqrt{\frac{1}{4}+\frac{\pi}{1-\pi}\frac{(z(hp,hu)-z(\ell p,\ell p))(z(hp,hu)-z(hu,hu))}{(z(hu,hu)-z(\ell u,\ell u))^2}}.\end{aligned}$$

A sufficient condition is

$$\left(\frac{z(hp,hp)-z(\ell p,\ell p)}{z(hp,hu)-z(\ell p,\ell p)}\right)^2 < 1+\frac{z(hp,hu)-z(hu,hu)}{z(hp,hu)-z(\ell p,\ell p)},$$

which holds if $z(hp,hu)$ is sufficiently close to $z(hp,hp)$.

The second case arises when $(1-\pi)e_u < \pi e_p$, that is, when

$$\frac{1-\pi}{\pi} < \frac{z(hp,hu)-z(\ell p,\ell p)}{z(hp,hu)-z(\ell u,\ell u)}.$$

Then $v^A(hu) = z(hp,hu)/2$ and

$$\begin{aligned}v^A(hp) =&\frac{z(hp,hp)-z(\ell p,\ell p)}{2}\\&-\frac{1-\pi}{\pi e_p}\frac{(z(hp,hu)-z(\ell u,\ell u))(z(z(hp,hp)-z(hp,hu))}{4}.\end{aligned}$$

This implies that

$$\begin{aligned}e_p^A =&\frac{z(hp,hp)-z(\ell p,\ell p)}{4}\\&+\frac{1}{2}\sqrt{\frac{(z(hp,hp)-z(\ell p,\ell p))^2}{4}-\frac{1-\pi}{\pi}(z(hp,hu)-z(\ell u,\ell u))(z(hp,hp)-z(hp,hu))}.\end{aligned}$$

Comparing surplus, $S^0 < S^A$ if

$$\begin{aligned}&\left(\frac{z(hu,hu)-z(\ell u,\ell u)}{z(hp,hu)-z(\ell u,\ell u)}\right)^2\\&< 1-\frac{z(hp,hp)-z(hp,hu)}{z(hp,hu)-z(\ell u,\ell u)}+\frac{\pi}{1-\pi}\left(\frac{z(hp,hp)-z(\ell p,\ell p)}{(z(hp,hu)-z(\ell u,\ell u))}\right)^2\\&\qquad\times\sqrt{\frac{1}{4}+\frac{1-\pi}{\pi}\frac{(z(hp,hu)-z(\ell u,\ell u))(z(hp,hp)-z(hp,hu))}{(z(hp,hp)-z(\ell p,\ell p))^2}}.\end{aligned}$$

Again a sufficient condition is

$$\left(\frac{z(hu,hu)-z(\ell u,\ell u)}{z(hp,hu)-z(\ell u,\ell u)}\right)^2 < 1-\frac{z(hp,hp)-z(hp,hu)}{z(hp,hu)-z(\ell u,\ell u)},$$

which holds if $z(hp,hu)$ is sufficiently close to $z(hp,hp)$. □

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 premarital 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.

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 non-contractible 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.

Kaneko, M. and Wooders, M. H. (1986), 'The core of a game with a continuum of players and finite coalitions: the model and some results', *Mathematical Social Sciences* **12**, 105–137.

Kaneko, M. and Wooders, M. H. (1996), 'The nonemptiness of the f-core of a game without side payments', *International Journal of Game Theory* **25**, 245–258.

Kremer, M. and Lavy, D. (2008), 'Peer effects and alcohol use among collge students', *Journal of Economic Perspectives* **22**(3), 189–206.

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.

Lerner, J. and Malmendier, U. (2013), 'With a little help from my (random) friends: Success and failure in post-business school entrepreneurship', *Review of Financial Studies* **26**(10), 2411–2452.

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.

Stinebrickner, R. and Stinebrickner, T. R. (2006), 'What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds', *Journal of Public Economics* **90**(8-9), 1435–1454.

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