Discrimination and Assimilation*

Jon X Eguia†
New York University

August 27, 2013

Abstract

I present a theory of assimilation in a heterogeneous society composed of two groups with distinct social norms and unequal statuses. Members of the group with a relatively disadvantaged status face an incentive to assimilate, embracing the norms of the more advantaged group. The cost of assimilation is endogenous and strategically chosen by the advantaged group to screen those seeking to assimilate. In equilibrium, only highly skilled agents, who generate positive externalities, choose to assimilate. The theory provides a novel explanation of the so called "acting white" phenomenon, in which students from disadvantaged ethnic groups punish their co-ethnics who succeed academically. I show that punishing success and thus raising the cost of acquiring skills needed to assimilate is an optimal strategy by low ability students to keep their more able co-ethnics in the disadvantaged group.

JEL Codes: J15, D71, Z13, D62, I24.

Keywords: Discrimination. Assimilation. Acting White. Peer effects. Social norms.

*I am grateful to Will Terry for extensive comments, to Drew Conway for excellent research assistance, to Dan Bernhardt, Alberto Bisin, Renee Bowen, Oeindrila Dube, Raquel Fernandez, Loukas Karabarbounis, Rachel Kranton, Maggie Penn, Carlo Prato, Jakub Steiner, Leonard Wantchekon, participants at the 2012 NBER Political Economy Summer Institute and audiences at talks at MPSA 2010, LSE, Northwestern-Kellogg, Stanford-GSB, Chicago, Harvard/MIT, NYU, Oslo, USC, UPF, Bristol and Leicester for their suggestions, and to the Ford Center for Global Citizenship and the Center for Mathematical Studies at the Kellogg School of Management for financial support during the academic year 2010-11.

†Email: eguia@nyu.edu. Mail: 19 West 4th St, 2nd floor, Dept. Politics, NYU. New York, NY 10012.

"When in Rome, do as the Romans do" (St. Ambrose, bishop of Milan, 384 AD).

In a heterogeneous society divided along cultural or ethnic cleavages, in which one social group enjoys a greater status or position of privilege, members of relatively disadvantaged groups face an incentive to assimilate into the more advantaged group, adopting its social norms and culture. Discrimination against those who seek to assimilate makes assimilation more difficult. I address two intimately related questions: When is it optimal for members of disadvantaged groups to assimilate? What are the incentives for members of the advantaged group to be receptive or hostile toward assimilation?

I present a theory of assimilation in a society comprised of two groups of agents: those with an advantaged background, who are exogenously endowed with favorable status, social capital or wealth, and those with a disadvantaged background, who lack this endowment. Agents are characterized by their background and their ability. Agents generate externalities for members of the group to which they ultimately belong; agents endowed with more status or wealth and more skilled agents generate more positive externalities. Disadvantaged agents choose whether or not to assimilate by joining the advantaged group. Advantaged agents choose how difficult it is to assimilate and join their group.

I find that agents with an advantaged background optimally screen those who seek to assimilate by choosing a difficulty of assimilation such that the agents who assimilate are precisely those whose skills are sufficiently high so that they generate a positive externality to the group. Comparative statics show that the equilibrium difficulty of assimilation increases in the exogenous endowment gap between groups. I argue that in order to screen optimally so that only the more able individuals assimilate, acceptance into the advantaged group must be based on malleable individual traits and behaviors that correlate with ability, and not on immutable characteristics that are uncorrelated with talent, such as skin color or place of birth.

The theory provides a novel explanation of the "acting white" phenomenon. Acting white refers to the seemingly self-hurting behavior by African-American and Hispanic students in the US who punish their peers for achieving academic excellence. While white students' popularity and number of friends increases with grades, African-American and Hispanic students who obtain top grades are *less* popular than their co-ethnics with lower grades (Fryer and Torelli 2010).

The traditional explanation (Fordham and Ogbu 1986, Fordham 1996) is cultural: African-Americans embrace academic failure as part of their identity and shun those who defy this identity by studying, and the rationale for this defeatist identity was that society denied African-Americans career opportunities and did not reward their effort. McWhorter (2000) argues that African-Americans engage in self-sabotage: society would reward African-Americans if they made an effort to excel, but they convince themselves that effort is not rewarded, and thus they do not exert effort. However, neither of these accounts fits well with recent empirical findings (Fryer and Torelli 2010).

Austen-Smith and Fryer (2005) propose an alternative theory based on the opportunity cost of studying: students who are socially inept do not enjoy their leisure time, so they choose to study, while other students differentiate themselves from the socially inept by choosing not to study. While compelling, this reasoning applies to all ethnicities, and thus it cannot explain the asymmetry across ethnic groups which is the essence of the acting white phenomenon.

I present a theory that fits the empirical findings of Fryer and Torelli (2010) and explains why African-American and Hispanic students, but not white students, experience a negative correlation between popularity and high grades.

I show that in equilibrium, students in underprivileged social groups optimally punish their overachieving co-ethnics. The incentive to deter excellence affects *only* disadvantaged groups because disadvantaged overachievers acquire skills to assimilate into a more privileged social group. Since highly able individuals generate positive externalities for the group in which they end up, and since society makes assimilation too difficult for the less able disadvantaged students, the second best outcome for this latter group of students is to retain the more able co-ethnics in their community. They achieve this by punishing academic excellence in order to deter the more able students from acquiring the skills necessary to assimilate. If we define “white” as a set of socioeconomic and cultural traits and not as a color, we can say that black students punish their most able co-ethnics for acting white because acting white is a prologue to becoming white.

Beyond the specific case of explaining the acting white phenomenon, the broader theory is applicable to social settings in which an outsider such as an immigrant may assimilate and join mainstream society. An immigrant can choose to adapt as quickly and fully as possible to the local culture, language, food, music, sports and social norms; or the immigrant can settle in a distinctly ethnic neighborhood where the culture of the immigrant's motherland is strong, declining to absorb the values, norms and customs prevalent in the rest of society.[1]

The cost of assimilation depends crucially on the attitude of the members of the social group that the migrant seeks to join. Sniderman, Hagendoorn and Prior (2004) find that Dutch citizens favor immigration by highly educated workers, and not by those who are only suited for unskilled jobs. Hainmueller and Hiscox (2010) refine this finding, distinguishing not only which immigrants inspire more negative reactions, but also which citizens (rich or poor) are more favorable toward each set of immigrants. They find that rich and poor US citizens alike strongly prefer high-skilled immigration and are opposed to low-skilled immigration. The theory I present in this paper is fully consistent with these results: economic self-interest leads low-skilled and high-skilled citizens alike to only welcome assimilation by high-skilled agents.

This paper builds upon an extensive literature on theories of social identity formation.[2] The literature on the economics of culture argues that minorities adopt and pass on to their descendents identities that are anti-achievement (Akerlof and Kranton 2000), traditional (Bénabou and Tirole 2011) or ethnic (Bisin and Verdier 2000 and 2001) because if they shed this identity and embrace the productive/modern/majority identity, they suffer an exogenously given cost. Shayo (2009) and Klor and Shayo (2010) theorize that agents would like to identify with a high status group formed by agents similar to them.

Identity theories teach us that given a sufficiently high exogenous cost of assimilation, it is not optimal to assimilate. I propose a theory that recognizes that the difficulty of

[1] If first generation immigrants do not assimilate, later generations of individuals brought up in the culture of an ethnic minority and not in the predominant culture of their land of residence, such as Turks in Germany, or Hispanics and other minorities in the US. face a qualitatively similar choice.

[2] For interdisciplinary perspectives on identity, see the surveys by Hogg (2003) in social psychology; Hill (2007) in law and economics, and Jenkins (1996) in all the social sciences.

assimilation is endogenous: it depends on the actions of the agents with an advantaged background. The opportunities for friendship and social connections, and the externalities experienced by an agent depend less on her own identity (her concept of self) and more on how she behaves, on what other agents think of her, and on how they treat her as a result. Identity theories do not ask why agents with an advantaged background discriminate against those who seek to assimilate: I show that discrimination arises in equilibrium as agents pursue their own selfish interests.

Research that focuses on behavior and on social interactions more than on an internal notion of self seeks to identify conditions that lead agents to learn a common language (Lazear 1999), to form friendships (Currarini, Jackson and Pin 2009 and 2010; Fong and Isajiw 2000; Echenique, Fryer and Kaufman 2006; Patacchini and Zenou 2006; Marti and Zenou 2009), to go on dates (Fisman, Iyengar, Kamenica, and Simonson 2008) and to marry (Eeckhout 2006, Fryer 2007) across ethnicities and races.[3] As in this paper, the focus is on behavior and interactions with others, not on an introspective concept of self.

A closer reference is Fryer's (2007a) theory of endogenous group choice. Agents face an infinitely repeated choice to invest in skills that are useful only to a narrow group, or in skills that are valued by society at large. Members of the narrow group reward the accumulation of group specific skills by greater cooperation with the agent. Fryer's theory features multiple equilibria under standard folk theorem arguments. He describes one equilibrium in which agents invest in group-specific skills, but since other equilibria yield different (and outright contradictory) empirical implications, the model lacks predictive power. Whereas, I show that disadvantaged agents suffer pressure from their peers to acquire a lower level of human capital in all equilibria. My theory generates unambiguous empirical implications that are consistent with the previously poorly explained findings by Hainmueller and Hiscox (2010) on attitudes toward immigration, and Fryer and Torelli (2010) on the acting white phenomenon.

The rest of the paper is organized as follows. First, I present the theory of assimilation.

[3]Friendships, dates and marriages are all positive interactions. I study societies where the alternatives are assimilation and peaceful segregation. Societies where a more plausible alternative to assimilation is inter-ethnic conflict face a different strategic environment, discussed among others by Fearon and Laitin (2000) and Caselli and Coleman (2006).

Then, I introduce peer pressure to the theory, to explain the acting white phenomenon, and I discuss how this theory fits available evidence on acting white better than alternative explanations.

1 Theory

Consider a society with a continuum of agents of unit mass. Agents are distinguished by their background and their ability, both of which are exogenously given. The background of a half of the agents is advantaged. Let $\mathcal{A}$ denote the set of agents with an advantaged background. Each agent $i \in \mathcal{A}$ has an endowment $e_{\mathcal{A}} > 0$. The other half of the agents, denoted by $\mathcal{D}$, have a disadvantaged background, and their endowment is $e_{\mathcal{D}} = 0$. I interpret this exogenous endowment very broadly, to include both wealth and also less tangible assets such as status or social capital accumulated by members of the group. This difference in endowment captures whatever initial advantage there is to be born in $\mathcal{A}$ instead of in $\mathcal{D}$.

Let θ_i denote the exogenously given ability or talent of agent i. Individual ability is private information. Assume that for each set of agents $\mathcal{J} \in \{\mathcal{A},\mathcal{D}\}$ the distribution of ability over $\mathcal{J}$ is uniform in $[0, 1]$.

Agents choose their skill and their social group.

Let s_i be the skill of agent i. Skill is endogenous, strategically chosen by agent i, subject to the constraint that $s_i \in [0, \theta_i]$. An agent's innate ability is an upper bound on how skilled the agent can become.

Assume that there are two social groups A and D, characterized by two competing sets of social norms and actions expected from their members. Members of the advantaged social group A speak in a certain language, with a certain accent. They adhere to a dress code, body language and pattern of behavior in social situations, eat certain foods and not others, and spend their leisure time on specific activities. Assume that every agent with an advantaged background immediately belongs to the advantaged social group, that is, $\mathcal{A} \subseteq A$.

An alternative set of norms, behaviors and actions is characteristic of members of the second, disadvantaged social group D. I assume that there is nothing intrinsically better or worse about either set of actions and norms; their only relevant feature is that agents with an advantaged background grow up embracing the advantaged norms as their own, whereas, agents with a disadvantaged background are brought up according to the disadvantaged social norms.

Notice that I use calligraphic letters $\mathcal{J} \in \{\mathcal{A}, \mathcal{D}\}$ to refer to the exogenous partition of the set of agents according to their background, while the standard letters A and D refer to the partition of agents into social groups, which depends on the assimilation decisions, as follows.

I assume that while many agents from a disadvantaged background are firmly attached to the disadvantaged social group D and have no choice but to belong to it, a fraction $\lambda \in (0, 1]$ of agents from a disadvantaged background can choose whether or not to join the advantaged social group A. Let $\mathcal{D}_Y \subset \mathcal{D}$ denote this set of agents who choose their social group strategically and assume that the distribution of individual ability (θ_i) in $\mathcal{D}_Y$ is uniform in $[0, 1]$, the same as in $\mathcal{D}$ or $\mathcal{A}$. I interpret $\mathcal{D}_Y$ as the set of agents with a disadvantaged background who are not yet settled in life and have enough contact or exposure to agents with advantaged background to have an opportunity to observe these advantaged agents' behavior, internalize their norms and assimilate.[4] This paper is concerned with these agents' choice between joining social group D, or overcoming whatever hurdles they face to join the advantaged social group A.

The cost of assimilation

Any agent $i \in \mathcal{D}_Y$ can choose to belong to D at no cost, or she can learn how to follow the norms of the group A to then join A, but this learning is costly. Let $a_i \in \{0, 1\}$ be the choice of agent $i \in \mathcal{D}_Y$, where $a_i = 0$ denotes that i chooses to be part of group D and not to assimilate, and $a_i = 1$ denotes that agent i chooses to assimilate into the advantaged group A. Let a denote the decisions to assimilate by all agents in $\mathcal{D}_Y$. Formally, $a : [0, 1] \to \{0, 1\}$ is a mapping from ability to assimilation decision. Given a, the composition of the social groups is $A = \mathcal{A} \cup \{i \in \mathcal{D}_Y : a_i = 1\}$ and $D = \mathcal{D} \backslash \{i \in \mathcal{D}_Y : a_i = 1\}$.

[4] In the application of the theory to explain the acting white problem in subsection 1.2, I will interpret the set $\mathcal{D}_Y$ more precisely as the set of young agents with a disadvantaged background who attend desegregated schools.

The cost of assimilating is $a_i d c(s_i)$, where a_i acts as an indicator function making the cost zero if agent i does not assimilate; $d \geq 0$ is the difficulty of learning and embracing the patterns of behavior consistent with membership in A, and $c : [0,1] \longrightarrow \mathbb{R}_{++}$ is a continuously differentiable, strictly decreasing function, which captures the intuition that more skilled agents can adapt at a lower cost. Let C be the set of all such functions.[5]

The difficulty of assimilation d is endogenous. It can be interpreted as the level of discrimination: If advantaged agents are welcoming to those who assimilate, d is small. If the set of agents $\mathcal{A}$ is hostile to those who do not master the cultural prerequisites of membership in A, then d is high. Formally, I assume that an exogenously given finite subset $\mathcal{A}_F \subset \mathcal{A}$ of size N of agents with an advantaged background chooses d.[6] Size N can be as small as one, or arbitrarily large. Label these agents according to their ability, so that $\theta_1 < \theta_2 < ... < \theta_N$. Each $i \in \mathcal{A}_F$ strategically chooses $d_i \in \mathbb{R}_+$, and the vector $(d_1, ..., d_N)$ aggregates into a difficulty of assimilation $d \in \mathbb{R}_+$. I do not specify exactly how this aggregation takes place: it could be that the discrimination/difficulty faced by those who assimilate is the minimum of all the individual d_i values, or the maximum, or the median, or any other order-statistic. I assume that for some integer $n \in \{1, ..., N\}$, d is the maximum real number such that at least n agents in $\mathcal{A}_F$ choose $d_i \geq d$. The intuition is that at least n agents must wish to erect a given barrier to assimilation in order for this barrier to materialize.[7]

Utility function

Agents derive utility from their endowment, from their skill, and from the externalities generated by the average endowment and skill of other agents in their social group. Let $\psi(e_i, s_i)$ be the direct utility that agent i obtains from her exogenous endowment and her own skill. The only assumptions on $\psi(e_i, s_i)$ are that it is continuous and strictly increasing in both arguments.

I assume that agents do not have others-regarding preferences, but there are externalities

[5] If we assume instead that the cost is a function of both ability and skill, $c(s_i, a_i)$, results are robust as long as this function is strictly decreasing in both terms.

[6] We could let all agents in $\mathcal{A}$ be involved in choosing d, but with an infinite number of agents, the strategic incentives to choose optimally vanish. Keeping the number finite generates strict incentives to choose optimally.

[7] The theory is robust if we assume instead that $d = \frac{1}{N} \sum_{i=1}^{N} d_i$.

or spillover effects among agents who belong to the same group. The externalities occur when agents who have more in common and take similar actions, interact with each other. Leisure and job opportunities, friendships, private and professional relationships develop more readily among agents who follow the same norms and take part in the same activities.[8] Agents with greater exogenous endowment e_i and with greater skill s_i generate more positive externalities to their friends and members of their group.

Formally, let e_A be the average endowment of agents in social group A. Note that $e_A \in \left[\frac{e_{\mathcal{A}}}{1+\lambda}, e_{\mathcal{A}}\right]$, where the lower bound is achieved if every $i \in \mathcal{D}_Y$ assimilates, and the upper bound is achieved if none assimilate. The average endowment of agents in D is in any case $e_D = e_{\mathcal{D}} = 0$. For any $J \in \{A, D\}$, let s_J be the average skill of agents in J.

Let $v(s_i, e_J, s_J)$ be the utility that an agent with skill s_i in social group $J \in \{A, D\}$ derives from the externalities coming from other agents in her social group when the average endowment and skill of these agents are e_J and s_J. Then, any $i \in \mathcal{A}$ (who by assumption belongs to A) and any agent $i \in \mathcal{D}_Y$ who assimilates receive utility from externalities $v(s_i, e_A, s_A)$, whereas agents with a disadvantaged background who do not assimilate receive utility from externalities $v(s_i, e_D, s_D)$.

Let $U(e_i, s_i, d, a)$ denote the utility function of agent i as a function of her own endowment and skill, the discrimination level d, and the assimilation decisions of all agents in $\mathcal{D}_Y$. If we let s_{-i} denote the skill of every other agent but i, and we let a_i be exogenously fixed at 0 for any $i \in \mathcal{A} \cup \mathcal{D} \backslash \mathcal{D}_Y$, the utility of an agent i in social group $J \in \{A, D\}$ can be written as:

$$U(e_i, s_i, s_{-i}, d, a) = \underbrace{\psi(e_i, s_i)}_{\textit{Direct Ut.}} + \underbrace{v(s_i, e_J, s_J)}_{\textit{Ut. Externalities}} - \underbrace{a_i dc(s_i)}_{\textit{Assim. cost}}. \tag{1}$$

Every agent enjoys the direct utility from her own endowment and skill, and the externalities from the average endowment and average skill of the social group they join; whereas, only young agents with a disadvantaged background who assimilate ($i \in \mathcal{D}_Y$ such that $a_i = 1$) incur the cost of assimilation $dc(s_i)$.

I assume that v is twice continuously differentiable, weakly increasing in s_i and strictly

[8] For recent experimental evidence on the economic benefits of social interaction, see Feigenberg, Field and Pande (forthcoming).

increasing in e_J and s_J. For $x, y \in \{s_i, e_J, s_J\}$, let v_{xy} denote the cross-partial derivative with respect to x and y. I assume that $v_{e_J e_J} \leq 0$ and $v_{s_J s_J} \leq 0$ (the marginal utility of externalities from average endowment and average skill is not increasing); $v_{e_J s_J} \geq 0$ (there is a complementarity between average group endowment and average group skill); and $v_{s_i e_J} = 0$ (every member of a group equally enjoys the externality from the group's average endowment).

Timing

I model the interaction of the agents as a game with three stages.

First, each agent in $\mathcal{A}_F \subset \mathcal{A}$ chooses her optimal discrimination level d_i. These choices aggregate into a difficulty of assimilation d, which becomes common knowledge.

Second, each agent chooses her skill $s_i \in [0, \theta_i]$. Skill, just like ability, remains private information. I assume in this section that acquiring skill up to the limit set by individual ability is costless, hence it is a dominant strategy for every agent to choose $s_i = \theta_i$. I relax this assumption in the next section to explain the acting white phenomenon.

Third, each agent $i \in \mathcal{D}_Y$ chooses whether or not to assimilate, $a_i \in \{0, 1\}$. These choices determine the average skill and endowment of each social group, and hence payoffs.

2 Results

I solve by backward induction, finding perfect Bayes Nash equilibria.

Given d, and given any strategy profile by all other members of $\mathcal{D}_Y$, an agent $i \in \mathcal{D}_Y$ prefers to assimilate only if her skill s_i is high enough so that her cost of assimilating $c(s_i)$ is sufficiently small. It follows that for any d, there is a cutoff $s(d)$ in the level of skill such that any member of $\mathcal{D}_Y$ chooses to assimilate if and only if her skill is above $s(d)$.

For any skill $s \in (0, 1)$, let $d(s)$ be the degree of difficulty of assimilation that makes s become this cutoff, so that only agents with skill above s choose to assimilate. I show that $d(s)$ is a function, not a correspondence, and I find two alternative sufficient conditions so that it is strictly increasing. If $d(s)$ is strictly increasing, $s(d)$ is a function and we obtain a unique solution. Each $i \in \mathcal{A}_F$ chooses $d_i^* = d(s_i^*)$ such that

$$s_i^* = \arg\max_{\{s\}} v(s_i, e_A(s), s_A(s)) \text{ s.t. } s_A(s) = \frac{1+\lambda-\lambda s^2}{2+2\lambda(1-s)} \text{ and } e_A(s) = \frac{e_{\mathcal{A}}}{1+\lambda(1-s)},$$

where $e_A(s)$ and $s_A(s)$ are the average endowment and skill of the agents in A as a function of s given that agents in $\mathcal{D}_Y$ assimilate if and only if their skill is above s. Because the rule that aggregates the chosen vector of d_i^* for each $i \in \mathcal{A}_F$ into d^* is strategy-proof, it is dominated for any i to choose any d_i other than the one that would maximize her own utility.

The first result below states that under either of two sufficient conditions (neither of which is necessary), there exists a unique equilibrium, and in this equilibrium agents assimilate if and only if their ability is sufficiently high. The result holds given any functional form of the direct utility ψ and utility from externalities v that satisfy the stated assumptions.

Proposition 1 *For any cost function $c \in C$, there exists $\lambda_c \in (0,1]$ such that if $\lambda \leq \lambda_c$, then i), ii) and iii) below hold. For any $\lambda \in (0,1]$, there exists $c_\lambda \in \mathbb{R}$, such that if $\frac{c'(s_i)}{c(s_i)} < c_\lambda$ for any $s_i \in [0,1]$, then i), ii) and iii) hold.*

i) There exist a unique perfect Bayesian equilibrium and a cutoff $\theta^* \in \left(\frac{1}{2}, 1\right]$ such that in this unique equilibrium, agents with a disadvantaged background assimilate if and only if their ability is above θ^*.

ii) There exists $\bar{e} \in \mathbb{R}_{++}$ such that if the difference in endowment $e_{\mathcal{A}} - e_{\mathcal{D}}$ is strictly less than $\bar{e}$, then $\theta^* < 1$ so that some agents assimilate.

iii) If in addition, individual skill and group skill are complements (the cross partial derivative $v_{s_i s_J}$ is non-negative), then in this equilibrium low-skilled agents with an advantaged background discriminate more than high-skilled agents, i.e. $d_1^* \geq d_2^* \geq ... \geq d_{N-1}^* \geq d_N^*$.

An equivalent restatement of Proposition 1 and its proof, along with all other proofs, are in the Appendix.

The intuition is that under either of the two sufficient conditions, agents with an advantaged background are able to optimally screen those who assimilate and join their group, by setting a positive but not too large difficulty of assimilation so that only agents with high ability (who in equilibrium are highly skilled) assimilate.

The first of the two sufficient conditions is that the size λ of the set of agents with a disadvantaged background who may assimilate is not too large. If this set is small, the assimilation decisions of other agents do not change the average skill or endowment of either group much, and each agent's assimilation decision depends mostly on her own ability: highly able agents become highly skilled and assimilate, less able agents find it too costly and do not assimilate.

A second sufficient condition is to assume that the cost of assimilation drops very rapidly (in relative terms) with skill, that is, that the derivative of the cost is very negative, relative to the magnitude of the cost, which implies that the cost faced by a more skilled agent is only a small fraction of the cost paid by a less skilled agent. If agents with unequal ability face such different incentives, the equilibrium is unique separating agents with ability above or below the cutoff, regardless of the size λ of the set of agents who can assimilate.

If agents with greater individual skill care more about their group's average skill, then agents with an advantaged background disagree on the optimal level of discrimination: highly skilled individuals, who appreciate their group's average skill more than less skilled individuals ($v_{s_i s_J} \geq 0$), want to discriminate less (strictly less if $v_{s_i s_J} > 0$) to assimilate more highly skilled agents with a disadvantaged background. Less skilled agents, who do not care as much for the increase in average skill that comes with assimilation, resent the decrease in average endowment and prefer higher barriers to assimilation to let fewer agents assimilate. Only if the endowment gap is too large, all agents with an advantaged background agree that it is best to not let anyone assimilate. Otherwise the solution is interior, and the cutoff for assimilation maximizes the utility of one advantaged agent, the one who is pivotal in determining the level of discrimination.

It is not necessary for the uniqueness result to hold that any of the two sufficient conditions holds, but if neither holds so agents face more homogeneous costs and the set of agents who can assimilate is large, then for some functional forms a cascade may occur: once the most skilled agents with a disadvantaged background assimilate, the average skill among the agents remaining in the disadvantaged group may be so low that agents with intermediate skills face a greater incentive to assimilate as well. If so, advantaged agents are no longer able to optimally screen, and it can occur (examples are available from the author) that the advantaged agents set a very high d^* to forestall the cascading assimilation of too many agents, or there can be multiple equilibria depending on whether agents with a disadvantaged background coordinate to assimilate in very small or in very large numbers.

Discrimination by means of imposing a difficulty of assimilation $d^* > 0$ is a screening device that the advantaged agents use to separate high skilled from low skilled agents, without a need to observe the actual skill level of the agent who assimilates.[9] Theories of statistical discrimination show that the inability to observe individual skill causes firms (Moro and Norman 2004) or a social planner (Norman 2003) to misallocate high-skilled agents to unskilled jobs. In contrast, in this manuscript's theory, agents sort themselves into their preferred social group. Since each agent knows her own skill, in equilibrium no agent is misallocated.

I describe in the Appendix three generalizations to the theory: (1) distinguishing between costs of assimilation based on behavioral norms that individuals must learn, and costs based on immutable exogenous traits such as race; (2) allowing for intrinsic preferences either against ethnic diversity (homophyly), or in favor of ethnic diversity; and (3) discussing a symmetric model in which $\mathcal{A}$ and $\mathcal{D}$ are each endowed with a different kind of endowment, and in which assimilation and discrimination occur in both directions. I find that in order to provide optimal screening for self-interested agents with an advantaged background, discrimination must be based on malleable traits (culture, behavior, etc.) and not on immutable traits that do not correlate with ability (skin color, place of birth, etc.). The results in Proposition 1 are robust if we let payoffs directly increase or decrease in diversity, or if we consider assimilation in both directions.

Two other factors outside the model could affect the incentives to assimilate. The first is dynamic considerations. In an overlapping generations model, if the most able agents with a

[9] By imposing a cost of assimilation, agents with an advantaged background both *discriminate against* all agents with a disadvantaged background, and –in a more favorable sense of the word- they *discriminate among* agents with a disadvantaged background, by passively separating the most talented among them, who assimilate, from the rest.

disadvantaged background assimilate, the distribution of ability among the older generation is not identical across groups; rather, it favors the advantaged group. This asymmetry increases the incentives to assimilate, leading to higher levels of discrimination. Second, if the scarcity of highly skilled members in the disadvantaged group allows them to attain positions of leadership or other rewards within the group, then high skilled agents have a counter incentive to stay in the disadvantaged group.[10] Since this incentive to stay only holds if fewer high skilled agents join the disadvantaged group, it must still be that in equilibrium some of them assimilate.

The equilibrium prediction that discrimination arises to deter low skilled agents from assimilating is consistent with survey evidence on attitudes toward immigration. Poor and rich US voters alike prefer high skilled immigration to low skilled immigration, and in fact oppose the latter (Hainmueller and Hiscox 2010). Hainmueller and Hiscox argue that economic theories cannot explain this finding: "economic self-interest, at least as currently theorized, does not explain voter attitudes toward immigration." From their abstract:

> "The labor market competition model predicts that natives will be most opposed to immigrants who have skill levels similar to their own. We find instead that both low-skilled and highly skilled natives strongly prefer highly skilled immigrants over low-skilled immigrants, and this preference is not decreasing in natives' skill levels. The fiscal burden model anticipates that rich natives oppose low-skilled immigration more than poor natives [...]. We find instead that rich and poor natives are equally opposed to low-skilled immigration."

The theory in this paper leading to Proposition 1, provides an explanation based strictly on self-interest that fully accounts for these attitudes.[11] The theory is also consistent with immigration policies that offer a path to naturalization and assimilation for highly skilled

[10] This could be modelled in the current framework by assuming that v depends on the whole distribution of skills in group J, and not only on the average s_J.

[11] Heinmueller and Hiscox consider theories based on labor market competition (Becker 1957) and on the cost of providing public services (Hanson, Scheve, and Slaughter 2007). Other theories of economic self-interest can also explain their findings. For instance, if high-skilled immigrants are net contributors to public finances, and low-skilled immigrants are a net burden, all natives may welcome high-skilled immigration and oppose low-skilled immigration for fiscal reasons (see Storesletten 2000).

immigrants (such as the "green card" in the U.S. or the "blue card" in the E.U.), while they keep the bulk of low skilled immigrants as undocumented or temporary "guest worker" aliens. As theorized, highly skilled immigrants are welcome to join society, whereas low skilled immigrants are not welcome to participate in civil society even when their labor is used as a production factor in the economy.

To study the comparative statics with changes in the endowment gap between groups, I relax the normalization that $e_{\mathcal{D}} = 0$, assuming instead that $0 \leq e_{\mathcal{D}} \leq e_{\mathcal{A}}$, so that I can study the effect of increases in the endowment of each group independently. Even if the gap remains the same, if the disadvantaged group becomes richer, the equilibrium level of difficulty of assimilation d^* decreases, and the proportion of agents who assimilate increases.

Proposition 2 *For any $c \in C$, there exist $\Delta \in \mathbb{R}_+$ and $\lambda_c \in (0,1]$ such that for any endowment gap $e_{\mathcal{A}} - e_{\mathcal{D}} \in (0, \Delta]$ and any $\lambda \leq \lambda_c$,*

i) The equilibrium difficulty d^* and cutoff for assimilation θ^* strictly decrease if $e_{\mathcal{D}}$ increases while $e_{\mathcal{A}}$ remains constant, and

ii) The equilibrium difficulty d^* and cutoff for assimilation θ^* decrease if both e_D and e_A increase in the same amount.

Furthermore, for any $\lambda \in [0,1]$, there exist $\Delta \in \mathbb{R}_+$ and $c_\lambda \in \mathbb{R}$ such that if $e_{\mathcal{A}} - e_{\mathcal{D}} \in (0, \Delta]$ and $\frac{c'(s_i)}{c(s_i)} < c_\lambda$ for any $s_i \in [0,1]$, then $i)$ and $ii)$ hold.

Result $i)$ says that if the endowment gap is not too large, assimilation increases as the endowment gap narrows. Result $ii)$ notes that assimilation also increases if both groups become richer, keeping the endowment gap constant. Both hold under either of the two sufficient conditions for equilibrium uniqueness.

The theory predicts that greater economic inequality across ethnic groups leads to less assimilation. This empirical implication can be tested using data on inequality across ethnic groups and on intermarriages.

Welfare analysis is not straightforward. Agents with different backgrounds have conflicting interests: agents with an advantaged background want the most skilled among the agents with a disadvantaged background to assimilate, but this assimilation makes the other agents with a disadvantaged background worse off. In equilibrium, and compared to the benchmark with no assimilation, agents with an advantaged background and the most able among those who assimilate benefit from assimilation, while agents with a disadvantaged background who do not assimilate become worse off.

In the next section I explain how agents with a disadvantaged background and low ability, who are harmed by the assimilation process we have described, react to protect their self-interest by raising the costs of exiting the disadvantaged social group. This self-interested reaction, strategically erecting barriers to exit, explains the acting white phenomenon.

3 Application: Acting White

"Acting white" is "a set of social interactions in which minority adolescents who get good grades in school enjoy less social popularity than white students who do well academically" (Fryer 2006). Fryer (2006) shows that "the popularity of white students increases as their grades increase. For black and Hispanic students, there is a drop-off in popularity for those with higher GPAs." This peer pressure against academic achievement leads minority adolescents to underperform, and contributes to the achievement gap of African-American and Hispanic students relative to white students.

I interpret the choice of a skill level $s_i \in [0, \theta_i]$ as the choice to attain a level of success in school. Students who choose $s_i < \theta_i$ do not achieve their potential, come out of school with fewer skills, and are less able to succeed in society. All else equal, every i prefers the highest possible skill s_i to maximize the direct utility $\psi(e_i, s_i)$. But all else is not equal: in some schools, peers may punish those who excel.

I introduce peer pressure into the theory. Recall that the set $\mathcal{D}_Y$ comprises the fraction λ of agents with a disadvantaged background who choose their social group strategically. Think of them as young minority students who attend desegregated high schools. Assume that these agents are susceptible to peer pressure. For symmetry in primitives, assume as well that a set $\mathcal{A}_Y \subset \mathcal{A}$ of size λ of young agents with an advantaged background are susceptible to peer pressure by other agents with an advantaged background.

I model peer pressure as follows: Let $l \in \mathcal{A}$ and $m \in \mathcal{D}$ be such that $\theta_i \leq \frac{1}{2}$ for $i \in \{m, l\}$. Agent l chooses a skill threshold $s_{\mathcal{A}}^P \in [0,1]$ and agent m chooses a skill threshold $s_{\mathcal{D}}^P \in [0,1]$. For $\mathcal{J} \in \{\mathcal{A}, \mathcal{D}\}$, threshold $s_{\mathcal{J}}^P$ is observed only by every $i \in \mathcal{J}_Y$. Every $i \in \mathcal{J}_Y$ who chooses $s_i > s_{\mathcal{J}}^P$ incurs a fixed cost $K > 0$. I let K be exogenously fixed at a strictly positive value for simplicity of exposition. Results hold if we endogenize K as follows: let l choose $K_{\mathcal{A}} \in [K_-, K^+]$ and m choose $K_{\mathcal{D}} \in [K_-, K^+]$ with $K_- < 0 < K^+$, and assume that for any $\mathcal{J} \in \{\mathcal{A}, \mathcal{D}\}$, any $i \in \mathcal{J}_Y$ who chooses $s_i > s_{\mathcal{J}}^P$ incurs punishment $K_{\mathcal{J}}$. Under this extension, in equilibrium m chooses $K_{\mathcal{D}} = K^+$ (proof available from the author). For simplicity, I directly assume $K_{\mathcal{A}} = K_{\mathcal{D}} = K^+ = K > 0$. I interpret this cost K as a reduced form that captures the social cost of overachieving in school, which may manifest itself in punishments as physical bullying, or more mildly, in the form of social disaffection.[12]

If we fix a_i at zero for any $i \in \mathcal{A}_Y$, for each background $\mathcal{J} \in \{\mathcal{A}, \mathcal{D}\}$, the utility function of an agent $i \in \mathcal{J}_Y$ in social group $J \in \{A, D\}$ can be written as:

$$\begin{aligned} \psi(e_i, s_i) + v(s_i, e_J, s_J) - a_i dc(s_i) - K \text{ if } s_i &> s_{\mathcal{J}}^P, \text{ and} \\ \psi(e_i, s_i) + v(s_i, e_J, s_J) - a_i dc(s_i) \text{ if } s_i &\leq s_{\mathcal{J}}^P, \end{aligned} \tag{2}$$

An agent $i \notin \mathcal{A}_Y \cup \mathcal{D}_Y$ is not susceptible so peer pressure and faces no assimilation decision. If she belongs to social group J, her utility function is, as in expression (1), $\psi(e_i, s_i) + v(s_i, e_J, s_J)$.

The timing is as follows:

1. Agent $m \in \mathcal{D}$ chooses a peer pressure threshold $s_{\mathcal{D}}^P$. Simultaneously, agent $l \in \mathcal{A}$ chooses the peer pressure threshold $s_{\mathcal{A}}^P$ and an arbitrary agent $h \in \mathcal{A}$ chooses the difficulty of assimilation d.[13]

[12] In practice, peer pressure must be implemented by a group. For simplicity, in this section I blackbox the collective implementation of peer pressure, assuming that the cutoff is chosen by a single individual, and the cost K incurred automatically. In a generalization (available from the author) I show that the equilibrium in Proposition 3 holds if punishments are determined by the aggregation of collective decisions.

[13] The main result is robust to variations in the timing of moves, such as letting d be chosen before $s_{\mathcal{D}}^P$ and $s_{\mathcal{A}}^P$ (as in earlier versions of the paper). I let only one agent determine d for ease of exposition: results hold if we let d be determined as the average, or as an order statistic of the list of d_i chosen by each i in a finite

2. Each agent i chooses her skill $s_i \in [0, \theta_i]$.

3. Agents in $\mathcal{D}_Y$ choose whether to assimilate or not. Payoffs accrue.

I solve by backward induction. First I explain the intuition, then I state the result.

Step 3 is solved as in the previous section, but now the distribution of skill in $\mathcal{A}$ and $\mathcal{D}$ may not be the same.

At step 2, any agent $i \notin \mathcal{A}_Y \cup \mathcal{D}_Y$ chooses skill $s_i = \theta_i$.[14] Any $i \in \mathcal{A}_Y$ chooses $s_i \in \{\theta_i, s^P_{\mathcal{A}}\}$ and any $i \in \mathcal{D}_Y$ chooses $s_i \in \{\theta_i, s^P_{\mathcal{D}}\}$.

At step 1, agent $l \in \mathcal{A}$ has no incentive to punish any agent with her background, because a higher skill level for any $i \in \mathcal{A}$ generates positive externalities to all members of $\mathcal{A}$. Hence, in equilibrium, $s^P_{\mathcal{A}} = 1$.

Whereas, agent $m \in \mathcal{D}$ who chooses $s^P_{\mathcal{D}}$ has an incentive to lower the skill level of some agents to prevent them from assimilating. Let Ω be an arbitrary pair of distributions of levels of skill in $\mathcal{A}$ and $\mathcal{D}$. For any Ω, there is a threshold function increasing in d such that in equilibrium of the subgame that follows given (d, Ω), agents with disadvantaged background choose to assimilate if and only if their skill is above the threshold. In equilibrium, agents with low ability and a disadvantaged background are hurt by this assimilation process: they are left behind. Fixing $s^P_{\mathcal{D}}$ below the threshold of assimilation deters some agents in $\mathcal{D}_Y$ from acquiring a skill level above the threshold and thus from assimilating. The optimal peer pressure maximizes e_D by inducing as many highly able agents as possible to stay in the disadvantaged group D, while lowering their skill level only just as much as it is necessary to prevent them from assimilating. Hence in every equilibrium, $s^P_{\mathcal{D}} < 1$ and some agents with a disadvantaged background are deterred from overachieving.

The following proposition, stated separately for each of our two familiar sufficient conditions, says that for any value γ above one half, if the endowment gap is not too large, then in any equilibrium, highly able disadvantaged students are punished for acquiring skills above threshold γ, while highly able advantaged students are not.

group $\mathcal{A}_F$, as in the previous section.

[14] It is strictly dominated for these agents to choose $s_i < \theta_i$. We could assume directly that $s_i = \theta_i$ for any agent $i \notin \mathcal{A}_Y \cup \mathcal{D}_Y$ to let only young agents choose their skill. This assumption would be more consistent with the interpretation that s_i measures accumulation of skills in school, and it would not change any result.

Proposition 3 *For any $\gamma \in \left(\frac{1}{2}, 1\right]$ and any $c \in C$, there exist $e(\gamma) \in \mathbb{R}_{++}$, and $\lambda(\gamma, c) \in (0, 1]$ such that if $e_{\mathcal{A}} \leq e(\gamma)$ and $\lambda < \lambda(\gamma, c)$ then:*

i) An equilibrium in which $s_{\mathcal{D}}^P < \gamma$ and $s_A^P = 1$ exists.

ii) In any equilibrium $s_{\mathcal{D}}^P < \gamma$ and $s_A^P = 1$.

For any $\gamma \in \left(\frac{1}{2}, 1\right]$ and any $\lambda \in (0, 1]$, there exist $e(\gamma) \in \mathbb{R}_{++}$ and $c(\gamma, \lambda) \in \mathbb{R}$ such that if $e_{\mathcal{A}} \leq e(\gamma)$ and $\frac{c'(s_i)}{c(s_i)} < c(\gamma, \lambda)$ $\forall s_i \in [0, 1]$, then i) and ii) hold.

Proposition 1 had shown that the equilibrium without peer pressure leads to assimilation, which harms agents with low ability and a disadvantaged background. Proposition 3 shows that these doubly disadvantaged agents respond optimally by punishing success in school. In all equilibria, highly able agents with a disadvantaged background are pressured to underperform; whereas, agents with an advantaged background are not. This is the acting white phenomenon.

Notice that in equilibrium, d is lower in the game with peer pressure than in the game without it: fewer highly skilled agents assimilate, and as a consequence, the average skill level in A is lower, so intergroup differences are smaller, making assimilation less desirable. I illustrate these and other differences with a numerical example.

Example 4 *Let $e_{\mathcal{A}} = 4$, $\lambda = 0.1$, $c(s_i) = \frac{1}{s_i}$, $\psi(e_i, s_i) = e_i^{1/2} + 10s_i$, and $v(s_i, e_J, s_J) = e_J^{1/2} + 10s_J$. Let $U_{\mathcal{A}}$, $U_{\mathcal{D}-}$ and $U_{\mathcal{D}+}$ respectively denote the average utility of $\{i \in \mathcal{A}\}$, $\{i \in \mathcal{D}\}$ and $\{i \in \mathcal{D} : \theta_i \leq \frac{1}{2}\}$. Columns 2 and 3 in the table below compare the equilibrium outcomes under an assumption of no peer pressure ($K = 0$) in column 2, and peer pressure ($K = 1$) in column 3, where $\mathbf{s}_{\mathcal{D}}^P = 0.6$ and $s_{\mathcal{A}}^P = 1$ are part of the equilibrium.*

(1)	*(2) No peer pressure*	*(3) Peer pressure*	*(3)-(2)*
d^*	*1.341*	*1.314*	*-0.027*
θ^*	*0.610*	*0.676*	*0.066*
s_D	*0.487*	*0.488*	***+0.001***
$U_{\mathcal{A}}$	*14.077*	*14.074*	*-0.003*
$U_{\mathcal{D}}$	*9.896*	*9.800*	*-0.096*
$U_{\mathcal{D}-}$	*7.376*	*7.384*	***+0.008***

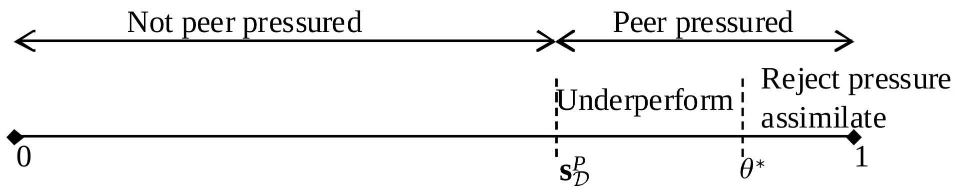

Figure 1: Acting white equilibrium with peer pressure.

While the acting white equilibrium makes all agents in $\mathcal{A}$ worse off and it reduces the average utility of agents in $\mathcal{D}$ and aggregate welfare, it makes agents with low ability and a disadvantaged background –the perpetrators of peer punishments– better off.

Figure 1 summarizes the effects of the acting white phenomenon on agents with a disadvantaged background. The horizontal axis measures ability. Students with ability below $s_{\mathcal{D}}^{P}$ are not subjected to any peer pressure. Students with ability above $s_{\mathcal{D}}^{P}$ are subjected to peer pressure to underperform (to acquire skills below their potential). Those with ability between the punishment threshold $s_{\mathcal{D}}^{P}$ and the equilibrium assimilation cutoff θ^* yield to the pressure and underperform to escape social punishments, while the most able reject the peer pressure, endure the consequent alienation from their co-ethnics, and ultimately assimilate into the advantaged community.

3.1 Discussion, Evidence and Policy Implications

I have presented a game-theoretic explanation of the acting white phenomenon: students in under-privileged communities dissuade their co-ethnics from acquiring skills in order to increase the cost of assimilation and deter exit from the community. This explanation has distinct empirical implications from those of alternative explanations in the literature (see the survey by Sohn 2011).

The "oppositional culture" theory of Fordham and Ogbu (1986) and Fordham (1996) posits that academic failure is an integral component of African-American group identity: whites embrace values of studiousness and hard work, while minorities reject these values, embracing instead a counterculture defined in opposition to the mainstream values, in particular in opposition to the pursuit of success at school. They find that students in the

1980s perceived activities such as speaking standard English, getting good grades, or going to libraries as distinctly "white" and they stress that to engage in these behaviors is to give up membership in the black social group. They trace back the roots of black students' self-identification with academic failure to a history of oppression in which whites (that is, society at large) negated their accomplishments regardless of their effort and objective merit.

Even if correct at the time, this account is anachronistic: the growing minority of African-American with stellar academic credentials who hold positions of leadership in society increasingly disprove the notion that recognition for intellectual achievements is a prerogative of whites. The Census data of 2000 notes that the average income of African-Americans with a high school, 2-year college, bachelor, master degree and professional degree is (respectively) 57%, 129%, 240%, 298% and 532% higher than the income of those who do not finish high school.[15] Academic success pays off for today's African-American students, even if Fordham and Ogbu (1986) are right and it formerly did not.

A second now traditional explanation is the "self-sabotage" argument posited by McWhorter (2000). The idea is that African-Americans engage in willful victimism, persuading themselves that discrimination in the job market is so pervasive that it makes costly accumulation of human capital not worthwhile. To the extent that self-saboteurs are deemed unworthy of social assistance, the term "self-sabotage" has normative consequences, and yet the term is misleading because it improperly anthropomorphizes the African-American minority: no individual African-American engages in self-sabotage; rather, students who have no ability to excel academically sabotage those who can excel.

An increasingly powerful argument against the sabotage explanation is that African-American attitudes have evolved away from the victimism decried by McWhorter (2000). Since the year 2000 a growing majority of African-Americans say that "blacks who cannot get ahead in this country are responsible for their own situation" and only a minority hold that discrimination is the main reason (Pew Research Center 2010).

The oppositional culture and the sabotage theories imply that the acting white problem ought to be more severe in schools with the least socioeconomic opportunities for upward

[15] In fact, the return for accumulation of cognitive skills is *greater* for African-Americans than for their white counterparts (Neal 2006).

mobility. The screening theory I have presented in this paper has the opposite empirical implication: the acting white phenomenon and the social price paid by the minority students who insist on achieving academic success should increase with the opportunities for upward mobility faced by the students.

Miron and Lauria (1998), Tyson (2006), Fryer (2006) and Fryer and Torelli (2010) test this implication. They all find that the acting white problem is more severe in less segregated (that is, in more racially integrated) schools: in predominantly black schools, which are those with the least opportunities for social mobility, "there is no evidence at all that getting good grades adversely affects students' popularity" (Fryer 2006). Fryer and Torelli (2010) find this "surprising." The screening theory offers an explanation: only black students in mixed schools are exposed to interaction with white students, so these students –as opposed to those in segregated schools- have greater opportunities to join a predominantly white social network, effectively abandoning the black community. In a fully segregated school, fears that a top student might shun the black community are minimized, as there is no alternative community that the student can join, so the acting white phenomenon does not occur. Fryer (2006) conjectures that perhaps the problem is attenuated if school desegregation leads to cross-ethnic friendships. The screening theory suggests the opposite: the greater the influence of white culture over black students, the greater the risk that the best black students assimilate. Fryer (2006) reports that indeed, greater inter-ethnic integration leads to a more severe acting white problem.

Summarizing the merits of the oppositional culture explanation and the sabotage theories, Fryer and Torelli (2005) note that these models "directly contradict the data in fundamental ways." Austen-Smith and Fryer (2005) propose an alternative explanation: high-school students shun studious colleagues because studiousness signals social ineptitude. Specifically, devoting time to study signals that the opportunity cost of time not spent in leisure is low because the individual is bad at leisure. While their argument is compelling, it applies to all races and social groups: their theory can explain why students do not want studious friends, but it cannot explain why only African-American and Hispanic students, and not non-Hispanic white students, exhibit this preference.

The asymmetry across ethnic groups is the essence of the acting white phenomenon. In the screening theory I have developed, this asymmetry is obtained as a main result (Proposition 3), derived from primitives (agents' utility functions, distribution of ability and technology for peer pressure) that are symmetric across groups, with the exception of an exogenous endowment. Solely from an unequal endowment, it follows that agents with a disadvantaged background discourage their peers's acquisition of skills, while agents with an advantaged background do not.

The signaling theory by Austen-Smith and Fryer (2005) and the screening theory in this paper disagree in one testable empirical implication. If students who obtain good grades are shunned because good grades signal social ineptitude, the popularity of a given student among students of *any* ethnicity must decrease with the student's grades. In particular, the popularity of African-American and Hispanic students among students of other ethnicities must decrease. If the screening theory is correct, minority students who obtain high grades are on a path away from their community and toward assimilation, which implies that while these students must be less popular among their co-ethnics (who will be left behind when the agent assimilates), they must be more popular among students outside her ethnicity (whom the agent is joining as she assimilates).

Fryer and Torelli (2010) test the relation between grades and out-of-race popularity measured as the number of friends of other races. They report (Table 5) that African-American or Hispanic students' out-of-race popularity increases in grades. Marti and Zenou (2009) report that in integrated schools (where the acting white phenomenon is more prevalent) "there are, mainly, two types of black students: those who have mostly white friends and those who choose mostly black friends" (see as well Patacchini and Zenou 2006). These findings together imply that African-American (and Hispanic) students with high grades have more white friends, while African-American (and Hispanic) students with lower grades build friendships mostly among their co-ethnics, which is fully consistent with the screening theory.

In summary, the screening theory of acting white fits well with the reported empirical findings on the greater prevalence of acting white in more integrated schools and the positive correlation between grades and out-of-race popularity, which clash with the predictions of the oppositional identity (Fordham and Ogbu 1986), self-sabotage (McWhorter 2000) and signaling theories (Austen-Smith and Fryer 2005).

This positive fit between the predictions of the screening theory and recent empirical findings establishes that variables in the data correlate as predicted by the theory, but it does not establish that the theory's causal mechanism is correct. As in all other studies of acting white, a concern remains that causality could be reversed, if it is not higher grades that cause a reduction in non-white friends, but rather, it is having few non-white friends that causes higher GPA scores. The longitudinal National Study of Adolescent Health (Add Health) data set can be used to test the screening theory addressing concerns about reverse causation. The Add Health study surveyed 20,745 adolescents in 1995, and then contacted 15,000 of them again in 2001-02 (wave III) and 2008-09 (wave IV). The screening theory posits that minority students with high grades are less popular among their co-ethnics because those with good grades are more likely to leave their social group. Using GPA scores and social network data from 1995, controls such as school type (private, public, urban, rural) and parental education, and social network data from 2001-02 and 2008-09, in future research we can check if indeed minority students with higher grades in 1995 are more likely to have left their original social group by 2008.

The punishment of high achieving African-American and Hispanic students is only an instance of a broader social phenomenon. In groups as diverse as the Buraku outcasts in Japan, Italian immigrants in Boston, the Maori in New Zealand and the working class in Britain, high-achievers have suffered a negative externality from their peer group (see Fryer 2007a or Sohn 2011 for a discussion). Hoff and Sen (2006) report a strikingly similar problem in the context of informal insurance provided by extended families in the developing world: "If the kin group foresees that it will lose some of its most productive members as the economy opens up, it may take collective actions ex ante to erect exit barriers." I interpret the acting white phenomenon as one such exit barrier.[16]

The screening theory's external validity as an explanation not just of acting white, but

[16] Religious doctrines opposing inter-faith marriage can also be understood as exit-deterrance strategies.

of the broader phenomenon that underprivileged communities deter exit by making skill acquisition costly, is testable. Students in rural schools face an analogous strategic environment: academic success leads to migration to the city. Therefore, the theory predicts rural students who obtain top grades to be less popular, regardless of their race. In the United States, this can be tested using the Add Health dataset. An analogous prediction applies to other countries and contexts; in the words of Fryer and Torelli (2010): "any group presented with the same set of payoffs, strategies and so on, would behave identically."

The policy implications of the theory can be summarized in a single insight: create incentives so that students become stakeholders in the success of their most able classmates.

If the classmates of a very able student perceive it to be in their immediate interest that the student excels, they will see to it that they do not punish success. Coleman (1961) found in the 1950s that athletes were the most popular students, and argued that athletes are popular because their effort results in honor and glory for the whole school. Whereas, studying only produces an individual gain. There is little positive spillover for her classmates and neighbors if a high-school student from an underprivileged neighborhood succeeds in high school and moves away to start a new life in college.

Policy interventions that provide contingent rewards based on observed behavior can change individual incentives in the classroom setting. Slavin (2009) surveys international financial incentives schemes aimed to increase education achievements and finds that these schemes have positive results in developing countries, but not in developed countries. Under these schemes, individuals are rewarded for their own behavior or achievement (a student gets a cash amount if she attends class, or if she gets a given grade, etc.), without any attention to peer effects. These incentives reinforce the perception that educational achievement is a purely individualistic good.

I suggest instead to distribute the conditional rewards to a group of peers, and not to an individual. A program that rewards every classmate or peer of a good student changes educational achievement from an individualistic good that only benefits the student, into a public production good that immediately benefits every member of the community, by means of the contingent collective reward. I conjecture that under these incentives, the most able students who produce the public good enjoyed by all their classmates would no longer lose popularity for achieving the high grades that deliver these public goods.

4 Appendix

First I describe three generalizations to the model. A detailed formalization, and precise results with their proofs for these generalizations are available from the author.[17]

Following the description of these generalizations, I provide the proofs of the propositions contained in the theory section of the paper.

Different kinds of discrimination

In an ethnically divided society, agents in $\mathcal{D}$ may differ from those in $\mathcal{A}$ with respect to some immutable, exogenous characteristic such as skin color, beside their differences in malleable traits such as cultural patterns and their difference in the endowment. In principle, advantaged agents could choose to make assimilation more difficult by discriminating on the exogenous and immutable traits, on the endogenous and malleable traits, or on both.

These two types of discrimination are qualitatively different: Discrimination based on immutable traits imposes a lump sum cost on every agent who wishes to assimilate. Whereas, discrimination based on endogenous traits imposes a cost that is negatively correlated with the agent's ability to learn and acquire the required traits, making it possible to screen agents according to type. So, if agents with an advantaged background seek to harness the positive externalities provided by highly skilled individuals, an optimal discrimination policy must be based on an endogenous correlate of ability such as the ease of learning the arbitrary cultural norms of group A, rather than on an ascriptive characteristic that offers no information about the person's skills.

Put it differently, even if advantaged agents care only about their self-interest and are unconcerned about the welfare of disadvantaged agents, as long as they are strategic, they do not discriminate on the basis of immutable characteristics such on skin color, race, place of birth. Rather, strategic agents with an advantaged background prefer to screen on the

[17] All the material available from the author is also available at http://dl.dropbox.com/u/9574908/EguiaDnAJun2013Addfile.pdf

basis of some observable characteristic that correlates with ability and skill. Agents with an advantaged background can construct and use a set of norms that are less costly to acquire for highly skilled agents, and then they can adopt a simple cut-off rule: Agents with a disadvantaged background who acquire a sufficiently high proficiency in the set of norms of A must be very skilled, and thus they should be assimilated, while agents who do not acquire such ease with the chosen norms are rejected and not assimilated.

A qualification to this argument leads to the second generalization.

Intrinsic preferences for or against diversity

If agents have intrinsic preferences over exogenous attributes such as race or place of birth, they may prefer ceteris paribus to associate with those who look like them or come from the same town. The qualitative results in the theory are robust to these preferences: If agents in $\mathcal{A}$ are prejudiced or dislike some exogenous attribute of set $\mathcal{D}$, agents in $\mathcal{A}$ treat those in $\mathcal{D}$ as if the endowment gap was higher, and as a result the equilibrium difficulty of assimilation d^* rises and fewer agents assimilate. If agents in $\mathcal{D}$ dislike some exogenous attribute of $\mathcal{A}$, then agents in $\mathcal{D}$ act as if the endowment gap was smaller, and the equilibrium difficulty of assimilation d^* must be lower in order to entice agents with a disadvantaged background to assimilate. If both sets of agents dislike the exogenous attributes of the other set, then the effect on d is ambiguous, but the number of agents who assimilate is smaller, resulting in voluntary segregation. Whereas, if ceteris paribus diversity increases agents' payoffs, in equilibrium there is less discrimination and more assimilation.

A symmetric society

Consider a more symmetric strategic environment in which groups have different endowments that are not clearly ordered, and assimilation and discrimination occur in both directions. An interpretation of this symmetric version is that different agents have different priorities in life. Perhaps an economically disadvantaged group $\mathcal{D}$ enjoys a greater artistic or musical richness in its community. Members of $\mathcal{D}$ who care about traditional forms of wealth and have high ability seek to assimilate into the wealthier group A; and yet, at the same time, members of $\mathcal{A}$ who are not motivated by material possessions but experience a greater utility if they live in a community that is rich in arts and music may seek to assimilate into

D.

Let there be two classes of endowment, e and m. Every $i \in \mathcal{A}$ is endowed with e in quantity $e_{\mathcal{A}}$ and every $i \in \mathcal{D}$ is endowed with m in quantity $m_{\mathcal{D}}$, while $e_{\mathcal{D}} = m_{\mathcal{A}} = 0$. Every agent i who values wealth e behaves as in the benchmark model, so that if $i \in \mathcal{A}$, then i chooses to be a member of A at no cost, and if $i \in \mathcal{D}_Y$, then i assimilates if and only if s_i is sufficiently high. However, now assimilation goes both ways: Agent $i \in \mathcal{A}_Y$ who values m assimilates into D if and only if she is sufficiently skilled.

The main insight holds in this more symmetric environment: Each group wants only highly skilled agents to assimilate, and it imposes a positive level of discrimination or difficulty of assimilation to screen those who wish to assimilate.

Proofs of the results.

Proposition 1 *For any $c \in C$, there exist $\lambda_c \in (0,1]$ such that if $\lambda < \lambda_c$ then*

i) there exists a unique perfect Bayesian equilibrium, and a cutoff $\theta^* \in \left(\frac{1}{2}, 1\right]$, such that in this equilibrium, any $i \in D_Y$ with $\theta_i > \theta^*$ assimilates and any $i \in \mathcal{D}_Y$ with $\theta_i < \theta^*$ does not assimilate,

ii) there exists $\bar{e} \in \mathbb{R}_{++}$ such that *if $e_{\mathcal{A}} < \bar{e}$, then in this equilibrium $\theta^* < 1$, and*

iii) if $v_{s_i s_A} \geq 0$, then $0 \leq d_N^* \leq d_{N-1}^* \leq ... \leq d_2^* \leq d_1^*$.

Furthermore, for any $\lambda \in (0,1]$, there exists $c_\lambda \in \mathbb{R}_{++}$ such that if $-\frac{c'(s_i)}{c(s_i)} > c_\lambda$ for any $s_i \in [0,1]$, then i), ii) and iii) hold.

Proof. First step of the proof. At the second stage, observing d, each agent i chooses s_i. Since s_i is private information, the choice does not affect future play by any other agent, and since the utility for i is ceteris paribus higher with a higher s_i, it follows that it is strictly dominated for any agent to choose any $s_i \neq \theta_i$. Hence every i chooses $s_i = \theta_i$.

Second step: At the third stage, agents in $\mathcal{D}_Y$ choose whether or not to assimilate, given d and given the decisions on skill at the second stage. Eliminating strictly dominated strategies, every agent correctly believes that every other agent has chosen skill $s_i = \theta_i$.

Let $s_A(s)$ and $e_A(s)$ be the average skill and endowment in A and let $s_D(s)$ be the average skill of agents in D as a function of s assuming that agents in $\mathcal{D}_Y$ assimilate if and only if their type is above s. Then

$$\begin{aligned} e_A(s) &= \frac{e_{\mathcal{A}}}{1+\lambda(1-s)}, \\ s_A(s) &= \left[\frac{1}{2}+\lambda(1-s)\frac{1+s}{2}\right]\frac{1}{1+\lambda(1-s)} = \frac{1+\lambda-\lambda s^2}{2+2\lambda(1-s)}, \\ s_D(s) &= \left[s\frac{s}{2}+(1-\lambda)(1-s)\frac{1+s}{2}\right]\frac{1}{s+(1-\lambda)(1-s)} = \frac{1-\lambda+\lambda s^2}{2-2\lambda(1-s)}. \end{aligned} \tag{3}$$

Given any d and any strategy profile a_{-i} for every $j \in \mathcal{D}_Y\backslash\{i\}$, since $c(s_i)$ is strictly decreasing in s_i, agent i chooses $a_i = 1$ if and only if s_i is above some cutoff that depends on d and a_{-i}. For any $i, j \in \mathcal{D}_Y$ such that $s_i > s_j$, and given any d and any strategy profile $a_{-i,j}$ for every $h \in \mathcal{D}_Y\backslash\{i, j\}$, if i and j best respond, $a_j = 1$ implies $a_i = 1$. Hence, given any d, there exists a cutoff in $[0, 1]$ such that for any $i \in \mathcal{D}_Y$, $a_i = 1$ if and only if s_i is above the cutoff, which depends on d.

Let $d(s)$ be the value of d such that $i \in \mathcal{D}_Y$ with $s_i = s$ is indifferent between assimilating or not given that other agents assimilating if and only if their skill is above s. This value is unique.

Third step: I identify two conditions such that $d(s)$ is a strictly increasing function.

For any $x, y, z \in \mathbb{R}$, let $v(s_i, e_J, s_J)|_{s_i=x,e_J=y,s_J=z}$ denote the value of $v(s_i, e_J, s_J)$ evaluated at $s_i = x$, $e_J = y$ and $s_J = z$. Then

$$d(s) = \frac{v(s_i, e_J, s_J)|_{s_i=s,e_J=e_A(s),s_J=s_A(s)} - v(s_i, e_J, s_J)|_{s_i=s,e_J=0,s_J=s_D(s)}}{c(s)}.$$

Note that if $\lambda = 0$, then

$$d(s) = \frac{v(s_i, e_J, s_J)|_{s_i=s,e_J=e_{\mathcal{A}},s_J=\frac{1}{2}} - v(s_i, e_J, s_J)|_{s_i=s,e_J=0,s_J=\frac{1}{2}}}{c(s)},$$

which is a strictly increasing, continuously differentiable function, with

$$d'(s) = \frac{\begin{array}{c}\left[\frac{\partial}{\partial s_i} v(s_i, e_J, s_J)|_{s_i=s, e_J=e_{\mathcal{A}}, s_J=\frac{1}{2}} - \frac{\partial}{\partial s_i} v(s_i, e_J, s_J)|_{s_i=s, e_J=0, s_J=\frac{1}{2}}\right] c(s) \\ -c'(s)\left[v(s_i, e_J, s_J)|_{s_i=s, e_J=e_{\mathcal{A}}, s_J=\frac{1}{2}} - v(s_i, e_J, s_J)|_{s_i=s, e_J=0, s_J=\frac{1}{2}}\right]\end{array}}{[c(s)]^2} > 0.$$

For any $\lambda \in [0, 1)$, since $e_A(s)$, $s_A(s)$, $s_D(s)$, $c(s)$, $c'(s)$ are continuous in λ for any $\lambda \in [0, 1)$, $v(s_i, e_J, s_J)$ is continuous, and $c(s)$ is positive for any s, so both $d(s)$ and $d'(s)$ are continuous in λ for any $\lambda \in [0, 1)$. Therefore, there exists $\lambda_c > 0$ (which depends on v as well as c) such that if $\lambda < \lambda_c$, then $d'(s) > 0$.

Alternatively, for any $\lambda \in (0, 1]$,

$$\begin{aligned} c(s)d'(s) \quad = \quad & \left[\frac{d}{ds} v(s_i, e_J, s_J)|_{s_i=s, e_J=e_{\mathcal{A}}(s), s_J=s_A(s)} - \frac{d}{ds} v(s_i, e_J, s_J)|_{s_i=s, e_J=0, s_J=s_D(s)}\right] \\ & -\frac{c'(s)}{c(s)}\left[v(s_i, e_J, s_J)|_{s_i=s, e_J=e_{\mathcal{A}}(s), s_J=s_A(s)} - v(s_i, e_J, s_J)|_{s_i=s, e_J=0, s_J=s_D(s)}\right]. \end{aligned}$$

Since v is continuously differentiable, the first term in the subtraction on the right hand side is bounded. The expression in brackets in the second term is strictly positive. It follows that if $\frac{c'(s)}{c(s)}$ is sufficiently negative, $-\frac{c'(s)}{c(s)}$ is sufficiently positive so that the right hand side is strictly positive and thus $c(s)d'(s) > 0$ and hence $d'(s) > 0$.

Assume for the remainder of the proof that either λ is small or $\frac{c'(s)}{c(s)}$ is very negative, so that $d'(s) > 0$.

Fourth Step: Find the optimal d_i^* for each $i \in \mathcal{A}_F$.

$$\begin{aligned} \text{Let } s^*(s_i) \quad &= \quad \arg\max_{s \in [0,1]} v(s_i, e_J, s_J) \text{ s.t.} \\ e_J \quad &= \quad e_A(s) = \frac{e_{\mathcal{A}}}{1 + \lambda(1 - s)}, \\ s_J \quad &= \quad s_A(s) = \frac{1 + \lambda - \lambda s^2}{2 + 2\lambda(1 - s)}. \end{aligned} \tag{4}$$

Since $v(s_i, e_A(s), s_A(s))$ is continuous in s, it achieves a maximum on the compact set $[0, 1]$, so a solution exists. I show that for a sufficiently low $e_{\mathcal{A}}$, the solution must be interior.

First, $s = 0$ is not a solution, because $\frac{dv(s_i,e_A,s_A)}{ds} > 0$ at $s = 0$. Second, $s = 1$ is not a solution for a low enough $e_{\mathcal{A}}$, because if $s = 1$, then

$$\begin{aligned}
\frac{dv(s_i, e_A, s_A)}{ds} &= \lambda e_{\mathcal{A}} \frac{\partial v(s_i, e_A, s_A)}{\partial e_A} + \frac{-2\lambda(1+\lambda)+\lambda}{2} \frac{\partial v(s_i, e_A, s_A)}{\partial s_A} \\
&= \lambda e_{\mathcal{A}} \frac{\partial v(s_i, e_A, s_A)}{\partial e_A} + \frac{-\lambda-\lambda^2}{2} \frac{\partial v(s_i, e_A, s_A)}{\partial s_A}
\end{aligned}$$

which is negative if

$$e_{\mathcal{A}} < \frac{1+\lambda}{2} \frac{\frac{\partial v(s_i,e_A,s_A)}{\partial s_A}}{\frac{\partial v(s_i,e_A,s_A)}{\partial e_A}}.$$

Since the solution is interior, it satisfies the first order condition

$$\frac{dv(s_i, e_A, s_A)}{ds} = \frac{\partial e_A}{\partial s} \frac{\partial v(s_i, e_A, s_A)}{\partial e_A} + \frac{\partial s_A}{\partial s} \frac{\partial v(s_i, e_A, s_A)}{\partial s_A} = 0. \tag{5}$$

Note that

$$\frac{\partial e_A}{\partial s} = \frac{\lambda e_{\mathcal{A}}}{[1+\lambda(1-s)]^2} \text{ and } \frac{\partial s_A}{\partial s} = \frac{-2\lambda s[1+\lambda(1-s)]+\lambda(1+\lambda-\lambda s^2)}{2[1+\lambda(1-s)]^2},$$

so a solution $s = s^*(s_i)$ satisfies

$$\begin{aligned}
0 &= \frac{1}{[1+\lambda(1-s)]^2} \left(\lambda e_{\mathcal{A}} \frac{\partial v(s_i, e_A, s_A)}{\partial e_A} + \frac{-2\lambda s[1+\lambda(1-s)]+\lambda(1+\lambda-\lambda s^2)}{2} \frac{\partial v(s_i, e_A, s_A)}{\partial s_A} \right) \\
0 &= \lambda e_{\mathcal{A}} \frac{\partial v(s_i, e_A, s_A)}{\partial e_A} - \lambda \frac{(1+\lambda)(2s-1)-\lambda s^2}{2} \frac{\partial v(s_i, e_A, s_A)}{\partial s_A}. \qquad (6)
\end{aligned}$$

To show that $s^*(s_i)$ is a unique solution, I show that $\frac{d^2 v(s_i,e_A,s_A)}{ds^2} < 0$ for any $s > s^*(s_i)$. It is easily verified that total derivative of the right hand side of equation 6 is negative, that is:

$$\begin{aligned}
&\lambda e_{\mathcal{A}} \left(\frac{\partial^2 v(s_i, e_A, s_A)}{\partial^2 e_A} e'_A(s) + \frac{\partial^2 v(s_i, e_A, s_A)}{\partial e_A \partial s_A} s'_A(s) \right) - \lambda[(1+\lambda)-\lambda s] \frac{\partial v(s_i, e_A, s_A)}{\partial s_A} \\
&-\lambda \frac{(1+\lambda)(2s-1)-\lambda s^2}{2} \left(\frac{\partial^2 v(s_i, e_A, s_A)}{\partial e_A \partial s_A} e'_A(s) + \frac{\partial^2 v(s_i, e_A, s_A)}{\partial^2 s_A} s'_A(s) \right) < 0.
\end{aligned}$$

The first term inside the first parenthesis is negative because $v_{e_J e_J} < 0$ and $e'_A(s) > 0$ $\forall s \in [0,1]$ by assumption. The second term inside the parenthesis is negative because $v_{e_J s_J} \geq 0$ by assumption, and $s'_A(s)$ must be negative in order for equation 5 to hold. The second term in the subtraction is negative because the partial derivatives of $v(s_i, e_J, s_J)$ are positive. Expression $-\lambda\frac{(1+\lambda)(2s-1)-\lambda s^2}{2}$ is negative if equation 6 holds. So it suffices to show that the two terms inside the last parenthesis are positive. The first term is positive because $v_{e_J s_J}$ is positive by assumption and $e'_A(s) > 0$ $\forall s \in [0,1]$, and the second is positive because $v_{s_J s_J} < 0$ by assumption and $s'_A(s)$ must be negative in order for equation 5 to hold. Hence, $s^*(s_i)$ is unique.

It follows that the optimal d for any agent $i \in \mathcal{A}_F$ is $d_i^* = d(s^*(s_i))$. Note that $\frac{dv(s_i, \omega_A, s_A)}{ds} > 0$ for any $s \leq \frac{1}{2}$, hence in order to satisfy the first order condition, it must be that $s^*(s_i) > \frac{1}{2}$, and since it has already been established that the solution is interior, it follows $s^*(s_i) \in (\frac{1}{2}, 1)$ and $d_i^* = d(s^*(s_i)) > 0$ as claimed. Each $i \in \mathcal{A}_F$ optimizes at a different value. Take the derivative of $\frac{dv(s_i, e_A, s_A)}{ds}$ from equation 5 with respect to s_i. If $v_{s_i s_A} \geq 0$ we obtain

$$\frac{\partial s_A}{\partial s}\frac{\partial^2 v(s_i, e_A, s_A)}{\partial s_A \partial s_i} \leq 0$$

hence an agent $j \in \mathcal{A}_F$ with $s_j \geq s_i$ satisfies the first order equation 5 by setting $s^*(s_j) \leq s^*(s_i)$ and thus $d_j^* \leq d_i^*$.

Fifth step: For any $i \in \mathcal{A}_F$, assimilation of agents with skill below $s^*(s_i)$ is detrimental to i, and assimilation of agents with skill above $s^*(s_i)$ is beneficial, hence each i has single-peaked preferences over the actual cutoff s. Since we have established that $d(s)$ is strictly increasing, it follows that i also has single-peaked preferences over d. The aggregation rule that determines d as a function of the vector $(d_1, ..., d_{|\mathcal{A}_F|})$ is strategy-proof (Moulin 1980) hence it is weakly dominated for any agent $i \in \mathcal{A}_F$ to choose any d_i other than $d_i = d_i(s^*(s_i))$. This results in cutoff $s^* = s^*(s_i)$ which, as shown in step four, is an interior solution if e_A is sufficiently low. Since, as argued in step one, $s_i = \theta_i$ for any $i \in \mathcal{D}_Y$, the ability cutoff θ^* for assimilation is $\theta^* = s^* < 1$. ■

Proposition 2

Proof. Note that

$$\begin{aligned} e_A(s) &= \frac{e_{\mathcal{A}} + \lambda(1-s)e_{\mathcal{D}}}{1+\lambda(1-s)} \text{ and} &(7) \\ \frac{\partial e_A}{\partial s} &= \frac{-\lambda e_{\mathcal{D}}[1+\lambda(1-s)] + [e_{\mathcal{A}} + \lambda(1-s)e_{\mathcal{D}}]\lambda}{[1+\lambda(1-s)]^2} = \frac{\lambda(e_{\mathcal{A}} - e_{\mathcal{D}})}{[1+\lambda(1-s)]^2}, &(8) \end{aligned}$$

so the first order condition is

$$\frac{dv(s_i, e_A, s_A)}{ds} = \frac{\partial e_A}{\partial s}\frac{\partial v(s_i, e_A, s_A)}{\partial e_A} + \frac{\partial s_A}{\partial s}\frac{\partial v(s_i, e_A, s_A)}{\partial s_A} = 0,$$

which implies (compare to equation 6 in the proof of proposition 1):

$$0 = \lambda(e_{\mathcal{A}} - e_{\mathcal{D}})\frac{\partial v(s_i, e_A, s_A)}{\partial e_A} - \lambda\frac{(1+\lambda)(2s-1) - \lambda s^2}{2}\frac{\partial v(s_i, e_A, s_A)}{\partial s_A}. \tag{9}$$

Given a fixed $e_{\mathcal{A}}$, if $e_{\mathcal{D}}$ increases, the first term in equation 9 decreases; the second term must then increase for the equality to hold. The second term is decreasing in s, so for any $j \in \mathcal{A}_F$, $s^*(s_j, e_{\mathcal{A}}, e_{\mathcal{D}})$ is decreasing in $e_{\mathcal{D}}$. As shown in the proof of proposition 1 for the case $e_{\mathcal{D}} = 0$, if λ is sufficiently small, or if $\frac{c'(s_i)}{c(s_i)}$ is sufficiently negative, $d(s)$ is strictly increasing in s. Generalize the notation to let $d(s, e_{\mathcal{A}}, e_{\mathcal{D}})$ denote the level of difficulty that makes $i \in \mathcal{D}_Y$ with skill $s_i = s$ indifferent between assimilation or not, as a function of both endowment levels. Since $e_{\mathcal{D}} = 0$ was merely a normalization, if λ and $e_{\mathcal{A}} - e_{\mathcal{D}}$ are sufficiently small, by the same argument $d(s, e_{\mathcal{A}}, e_{\mathcal{D}})$ is increasing in s. For any $e_1 > e_0$, $d(s, e_{\mathcal{A}}, e_{\mathcal{D}})|_{e_{\mathcal{D}}=e_1} < d(s, e_{\mathcal{A}}, e_{\mathcal{D}})|_{e_{\mathcal{D}}=e_0}$ because, given a fixed $e_{\mathcal{A}}$, the incentive to assimilate is lower if $e_{\mathcal{D}}$ is higher. Thus,

$$d(s, e_{\mathcal{A}}, e_{\mathcal{D}})|_{s=s^*(s_j, e_{\mathcal{A}}, e_1), e_{\mathcal{D}}=e_1} < d(s, e_{\mathcal{A}}, e_{\mathcal{D}})|_{s=s^*(s_J, e_{\mathcal{A}}, e_0), e_{\mathcal{D}}=e_1} < d(s, e_{\mathcal{A}}, e_{\mathcal{D}})|_{s=s^*(s_j, e_{\mathcal{A}}, e_0), e_{\mathcal{D}}=e_0}$$

so $d_j^*(s_j, e_{\mathcal{A}}, e_{\mathcal{D}}) = d(s, e_{\mathcal{A}}, e_{\mathcal{D}})|_{s=s^*(s_j, e_{\mathcal{A}}, e_{\mathcal{D}})}$ is strictly decreasing in $e_{\mathcal{D}}$ for each $j \in \mathcal{A}_F$, and thus the equilibrium difficulty $d^*(e_{\mathcal{A}}, e_{\mathcal{D}})$ is strictly decreasing in $e_{\mathcal{D}}$.

Similarly, for the second part of the proposition, given any sufficiently small fixed endowment gap $e_{\mathcal{A}} - e_{\mathcal{D}}$, if $e_{\mathcal{A}}$ and $e_{\mathcal{D}}$ increase in the same quantity, then $\frac{\partial v(s_i, e_A, s_A)}{\partial e_A}$ decreases by assumption (strictly if $v_{e_J e_J} < 0$), so the first term of the summation in equation 9 decreases. The rest of the argument is analogous to the case in the first part of the proposition. ■

Proposition 3 holds for any given functional form of the utility function U that satisfies the stated assumptions on ψ and v.

Proposition 3 *For any $\gamma \in \left(\frac{1}{2}, 1\right]$ and any $c \in C$, there exist $e(\gamma) \in \mathbb{R}_{++}$, and $\lambda(\gamma, c) \in (0, 1]$ such that if $e_{\mathcal{A}} \leq e(\gamma)$ and $\lambda < \lambda(\gamma, c)$ then:*

i) An equilibrium in which $s^P_{\mathcal{D}} < \gamma$ and $s^P_A = 1$ exists.

ii) In any equilibrium $s^P_{\mathcal{D}} < \gamma$ and $s^P_A = 1$.

For any $\gamma \in \left(\frac{1}{2}, 1\right]$ and any $\lambda \in (0, 1]$, there exist $e(\gamma) \in \mathbb{R}_{++}$ and $c(\gamma, \lambda) \in \mathbb{R}$ such that if $e_{\mathcal{A}} \leq e(\gamma)$ and $\frac{c'(s_i)}{c(s_i)} < c(\gamma, \lambda)$ $\forall s_i \in [0, 1]$, then i) and ii) hold.

Proof. Part I: I first prove the existence claim.

Let $\Omega = (\Omega_{\mathcal{D}_Y}, \Omega_{\mathcal{A}_Y})$ be the pair of distributive functions of the skill in $\mathcal{D}_Y$ and $\mathcal{A}_Y$.

Note first that at the second stage, for any $i \in \mathcal{J}_Y$ and $\mathcal{J} \in \{\mathcal{A}, \mathcal{D}\}$ choosing any $s_i \notin \{\theta_i, s^P_{\mathcal{J}}\}$ is strictly dominated either by $s_i = \theta_i$ or by $s_i = s^P_{\mathcal{J}}$. Therefore, in equilibrium $s_i \in \{s_i, s^P_{\mathcal{J}}\}$, hence $\Omega_{\mathcal{J}_Y}$ has uniform density on $[0, s^P_{\mathcal{J}})$ and positive mass only at $s^P_{\mathcal{J}}$.

At the third stage, by an analogous argument as in the proof of Proposition 1, there is a cutoff $s(d, \Omega) \in [0, 1]$ such that agent $i \in \mathcal{D}_Y$ chooses $a_i = 1$ if $s_i > s(d, \Omega)$ and chooses $a_i = 0$ if $s_i < s(d, \Omega)$. Unlike in the proof of Proposition 1, the cutoff may not be unique; if it is not unique, pick the solution with the fewest agents assimilating.

At the second stage, in anticipation of the equilibrium in stage 3, any $i \notin \mathcal{D}_Y \cup \mathcal{A}_Y$ and any agent $i \in \mathcal{J}_Y$ with $\theta_i < s^P_{\mathcal{J}}$ uniquely best respond by choosing $s_i = \theta_i$. Any agent $i \in \mathcal{J}_Y$ with $\theta_i > s^P_{\mathcal{J}}$ faces a trade-off: choosing $s_i = \theta_i > s^P_{\mathcal{J}}$ she incurs a cost K, but she derives a benefit in terms of direct utility ψ and in terms of a reduced cost of assimilation (if the assimilates). The benefit of choosing $s_i = \theta_i$ is increasing in θ_i, while the cost is fixed at K. Thus, there is a cutoff $\theta(s^P_{\mathcal{A}})$ such that $\theta(s^P_{\mathcal{A}}) > s^P_{\mathcal{A}}$ and such that any $i \in \mathcal{A}_Y$ with $\theta_i > \theta(s^P_{\mathcal{A}})$ chooses $s_i = \theta_i$ and any $i \in \mathcal{A}_Y$ with $\theta_i < \theta(s^P_{\mathcal{A}})$ chooses $s_i = s^P_{\mathcal{A}}$; and there is a second cutoff $\theta(d, s^P_{\mathcal{D}})$ such that $\theta(d, s^P_{\mathcal{D}}) > s^P_{\mathcal{D}}$ and any $i \in \mathcal{D}_Y$ with $\theta_i > \theta(d, s^P_{\mathcal{D}})$ chooses $s_i = \theta_i$ and any $i \in \mathcal{D}_Y$ with $\theta_i < \theta(d, s^P_{\mathcal{D}})$ chooses $s_i = s^P_D$. Both of these cutoffs depend crucially on parameter K.

At the first stage, note first that in equilibrium $(s^P_{\mathcal{A}})^* = 1$. Choosing $s^P_{\mathcal{A}} < 1$ causes any $i \in \mathcal{A}$ with $\theta_i \in (s^P_{\mathcal{A}}, \theta(s^P_{\mathcal{A}}))$ to choose skill $s_i = s^P_{\mathcal{A}} < \theta_i$, which reduces s_A, making l strictly worse off. Define $f(d) : \mathbb{R}_+ \longrightarrow [0, 1]$ as a function such that, assuming $s^P_{\mathcal{D}} = f(d)$, and given equilibrium play in stages 2 and 3, an agent $i \in \mathcal{D}_Y$ with skill $s_i = f(d)$ is indifferent between assimilating or not. Note however that $f(d)$ is only well defined for an interval; for very low values of d, the indifferent agent is not guaranteed to exist as all may prefer to assimilate; for sufficiently high d, all agents prefer not to assimilate. Similarly define $g(d) : \mathbb{R}_+ \longrightarrow [0, 1]$ as a function such that given d, given $s^P_{\mathcal{D}} = f(d)$, and given equilibrium play in stages 2 and 3, an agent $i \in \mathcal{D}_Y$ with ability $\theta_i = g(d)$ is indifferent between choosing skill $s_i = g(d)$ and assimilating or choosing skill $f(d)$; this function is well defined only if $f(d)$ is well defined. The intuition is that agents with ability $\theta_i \in [f(d), g(d))$ choose skill $s_i = f(d)$ and do not assimilate, and agents with ability $\theta_i \geq g(d)$ choose skill $s_i = \theta_i$ and assimilate.

If λ is sufficiently small, or if $\frac{c'(s_i)}{c(s_i)}$ is sufficiently negative, for any d such that $f(d)$ is well defined and takes an interior value (strictly less than one), $f(d)$ is strictly increasing. If $e_{\mathcal{A}}$ is sufficiently low, agent $h \in \mathcal{A}$ who chooses d strictly prefers an agent with skill $s_i = \gamma$ to assimilate. Let the equilibrium d^* be any d such that $f(d)$ is well defined, agent l strictly prefers an agent $i \in \mathcal{D}_Y$ with skill $s_i = f(d)$ to not assimilate (which implies $f(d) < \gamma$), and an agent with skill $s_i = g(d)$ (or with skill $s_i = 1$ if $g(d)$ is not defined) to assimilate. This equilibrium value d^* leads to assimilation for agents with ability $\theta_i \geq g(d^*)$. In equilibrium, $(s^P_{\mathcal{D}})^* = f(d^*)$. I check that $(s^P_{\mathcal{D}})^*$, d^*, and $(s^P_{\mathcal{A}})^*)$ are mutual best responses.

If agent $m \in \mathcal{D}$ deviates to $s^P_{\mathcal{D}} < (s^P_{\mathcal{D}})^*$, all agents in D (including m) become worse off because those with ability $\theta_i \in [s^P_{\mathcal{D}}, \theta(d^*, s^P_{\mathcal{D}})]$ lower their skill from $\min\{\theta_i, (s^P_{\mathcal{D}})^*\}$ to $s^P_{\mathcal{D}}$ to avoid the punishment K, leading to a decrease in θ_D. If m deviates to $s^P_{\mathcal{D}} > (s^P_{\mathcal{D}})^*$, then agents with ability $\theta_i \in ((s^P_{\mathcal{D}})^*, s^P_{\mathcal{D}}]$ choose $s_i = \theta_i$ and assimilate, again reducing s_D. Choosing $d < d^*$ causes those with skill $f(d^*)$ to assimilate, which makes h strictly worse off. Choosing $d > d^*$ causes those with ability $g(d^*)$ to not assimilate (if $g(d^*)$ is well defined), which makes h worse off, or it has no effect (if $g(d^*)$ is not well defined). Therefore, $(s^P_{\mathcal{A}})^* = 1$, d^* and $(s^P_{\mathcal{D}})^* = f(d^*)$ are best responses.

Part II: Next I show that $s^P_{\mathcal{D}} < \gamma$ and $s^P_A = 1$ in all equilibria.

In any equilibrium, $s^P_{\mathcal{A}} = 1$. Suppose not. Then agent l who chooses $s^P_{\mathcal{A}}$ can deviate to $(s^P_{\mathcal{A}})' = 1$. Only agents in $\mathcal{A}_Y$ observe this deviation, so only they react to it. The reaction consists of an increase in s_i from $s_i = s^P_{\mathcal{A}} < 1$ to $s_i = \theta_i$ for any $i \in \mathcal{A}_Y$ with $\theta_i \in (s^P_{\mathcal{A}}, 1]$, which increases s_A. Thus agent l prefers to deviate, and thus $s^P_{\mathcal{A}} < 1$cannot be sustained in equilibrium, so that $s^P_{\mathcal{A}} = 1$ in any equilibrium. Next, I show that in equilibrium, $s^P_{\mathcal{D}} < \gamma$.

The third stage does not have a unique solution. If $\Omega_{\mathcal{D}_Y}$ has an interval with density zero, and the skills cutoff for assimilation is in this interval, any value in the interval serves as a cutoff. Let $s_1(d, \Omega)$ and $s_2(d, \Omega)$ denote the lowest and highest possible values of the cutoff. The solution (and not just the cutoff that represents it) for the third stage may not be unique. Suppose $s_1(d, \Omega) \in [s^P_{\mathcal{D}}, s^P_{\mathcal{D}} + \varepsilon_1]$ for some small $\varepsilon_1 \geq 0$. Suppose that every $i \in \mathcal{D}_Y$ with skill $s_i \in [s^P_{\mathcal{D}}, s_1(d, \Omega)]$ deviates to $a_i = 1$. Then the average skill of each group suffer a discrete change, and if this change increases $s_A - s_D$, it can make each of the deviators strictly prefer to persist in the deviation. In this case, a second equilibrium arises in which $s(d, \Omega) = s^P_{\mathcal{D}} - \varepsilon_2$ for some $\varepsilon_2 \geq 0$. Let $v^1_{i,J}$ and $v^2_{i,J}$ denote the utilities from externalities derived by agent i in group J in the first and second of these two equilibria. Note that $\underset{\lambda \longrightarrow 0}{lim} (v^2_{i,J} - v^1_{i,J}) = 0$, which implies $\underset{\lambda \longrightarrow 0}{lim} \varepsilon_1 + \varepsilon_2 = 0$. Hence the two stage-game equilibria become arbitrarily close if λ is sufficiently small. Alternatively, for a fixed λ, if $\frac{c(s'_i)}{c(s_i)} \longrightarrow -\infty$, then again $\varepsilon_1 + \varepsilon_2 \longrightarrow 0$. Therefore, while the third stage can have multiple equilibria, the equilibrium outcomes become arbitrarily close to each other if λ is sufficiently small, or $\frac{c(s'_i)}{c(s_i)}$ is sufficiently negative.

Assume $s^P_{\mathcal{A}} = 1$ and $s^P_{\mathcal{D}} \geq \gamma$ for some $\gamma > \frac{1}{2}$. If $e_{\mathcal{A}}$ is sufficiently small, there exists an open interval O around γ such that agent h strictly prefers agents with skill level contained in O to assimilate. Assume if λ is sufficiently small or $\frac{c'(s_i)}{c(s_i)}$ is sufficiently negative so that $d(s)$ is strictly increasing everywhere except (possibly) at $s = s^P_{\mathcal{D}}$. Then the equilibrium d^* must be such that agents with skill $s_i = \gamma$ assimilate. Otherwise, agent h who chooses d becomes better off deviating to a lower d to let agents with skill $s_i = \gamma$ assimilate. A change in d may lead agent to coordinate on a different equilibrium and result on a discontinuous jump on the set of agents who assimilate making it impossible for h to target her exact intended cutoff for assimilation; however, as shown above, this discontinuous jump is small enough such that the cutoff for assimilation remains below γ and within the open interval O, making h strictly better off after the deviation. But then, agent m can become better off deviating to lower $s_{\mathcal{D}}^{P}$ to the actual cutoff of assimilation in this equilibrium, so that after the deviation, some agents stop assimilating. Thus $s_{\mathcal{D}}^{P} \geq \gamma$ cannot be sustained in equilibrium; in any equilibrium, $s_{\mathcal{D}}^{P} < \gamma$. ■

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics 115*(3), 715–753.

Austen-Smith, D. and R. G. Fryer (2005). An Economic Analysis of Acting White. *Quarterly Journal of Economics 120*(2), 551–583.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago, U. Chicago Press.

Benjamin Feigenberg, E. F. and R. Pande (Forthcoming). Experimental Evidence from Microfinance. *Review of Economic Studies*, –.

Bisin, A. and T. Verdier (2000). Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits. *Quarterly Journal of Economics 115*(3), 955–988.

Bisin, A. and T. Verdier (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory 97*(2), 298–319.

Bénabou, R. and J. Tirole (2011). Identity, Morals and Taboos: Beliefs as Assets. *Quarterly Journal of Economics 126*(2), 805–855.

Caselli, F. and W. J. Coleman (2006). On the Theory of Ethnic Conflict. NBER wp 12125.

Coleman, J. S. (1961). *The Adolescent Society*. Free Press New York.

Currarini, S., M. O. Jackson, and P. Pin (2009). An Economic Model of Friendship: Homophily, Minorities and Segregation. *Econometrica 77*(4), 1003–1045.

Currarini, S., M. O. Jackson, and P. Pin (2010). Identifying Sources of Race-Based Choice and Change in High School Friendship Network Formation. *Proceedings of the National Academy of Science 107*, 4857–4861.

de Marti, J. and Y. Zenou (2009). Ethnic Identity and Social Distance. CEPR n. 7566.

Echenique, F., R. G. Fryer, and A. Kaufman (2006, may). Is School Segregation Good or Bad? *American Economic Review 96*(2), 265–269.

Eeckhout, J. (2006). Minorities and Endogenous Segregation. *Review of Economic Studies 73*(1), 31–53.

Fearon, J. D. and D. D. Laitin (2000). Violence and the Social Construction of Ethnic Identity. *International Organization 54*, 845–877.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2008). Racial Preferences in Dating. *Review of Economic Studies 75*(1), 117–132.

Fong, E. and W. W. Isajiw (2000). Determinants of Friendship Choices in Multiethnic Society. *Sociological Forum 15*(2), 26–32.

Fordham, S. (1996). *Blacked Out: Dilemmas of Race, Identity and the Success at Capital High*. Chicago, University of Chicago Press.

Fordham, S. and J. Ogbu (1986). Black Students' School Successes: Coping with the Burden of 'Acting White'. *The Urban Review 18*(3), 176–206.

Fryer, R. D. (2006). Acting White. *Education Next 6*(1), –.

Fryer, R. D. (2007a). A Model of Social Interactions and Endogenous Poverty Traps. *Rationality and Society 19*(3), 335–366.

Fryer, R. D. (2007b). Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century. *Journal of Economic Perspectives 21*(2), 71–90.

Fryer, R. D. and P. Torelli (2005, May). An Empirical Analysis of 'Acting White'. NBER wp 11334.

Fryer, R. D. and P. Torelli (2010). An Empirical Analysis of 'Acting White'. *Journal of Public Economics 94*(5-6), 380–396.

Hainmueller, J. and M. J. Hiscox (2010). Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survery Experiment. *American Political Science Review 104*(1), 61–84.

Hanson, G. H., K. F. Scheve, and M. J. Slaughter (2007). Public Finance and Individual Preferences over Globalization Strategies. *Economics and Politics 19*(1), 1–33.

Hill, C. A. (2007). The Law and Economics of Identity. *Queen's Law Journal 32*(2), 389–445.

Hoff, K. and A. Sen (2006). *Poverty Traps*, Chapter The Kin System as a Poverty Trap? S. Bowles, S. Durlauf, K. Hoff eds. Princeton U. Press.

Hogg, M. A. (2003). *Handbook of Self and Identity*, Chapter Social Identity. Guilford Press.

Jenkins, R. (1996). *Social Identity*. Routledge, London.

Klor, E. F. and M. Shayo (2010). Social Identity and Preferences over Redistribution. *Journal of Public Economics 94*, 269–278.

Lazear, E. P. (1999). Culture and Language. *Journal of Political Economy 107*(6), S95–S126.

McWhorter, J. H. (2000). *Losing the Race: Self-Sabotage in Black America*. New York, Free Press.

Mirón, L. F. and M. Lauria (1998). Student Voice as Agency: Resistence and Accommodation in Inner-City Schools. *Anthropology and Education Quarterly 29*(2), 189–213.

Moro, A. and P. Norman (2004). A General Equilibrium Model of Statistical Discrimination. *Journal of Economic Theory 114*(1), 1–30.

Moulin, H. (1980). On Strategy-proofness and Single Peakedness. *Public Choice 35*(4), 437–455.

Neal, D. (2006). *Handbook of the Economics of Education*, Chapter Why has Black-White Skill Convergence Stopped? Eds E. Hanushek and F. Welch. Elsevier.

Norman, P. (2003). Statistical Discrimination and Efficiency. *Review of Economic Studies 70*(3), 615–627.

Patacchini, E. and Y. Zenou (2006). Racial Identity and Education. IZA discussion paper 2046.

Pew Research Center (2010, Jan). A Year After Obama's Election. Blacks Upbeat about Black Progress, Prospects. Survey.

Shayo, M. (2009). A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review 103*(2), 147–174.

Slavin, R. E. (2010). Can Financial Incentives Enhance Educational Outcomes? *Educational Research Review 5*, 68–80.

Sniderman, P. M., L. Hagendoorn, and M. Prior (2004). Predisposing Factors and Situational Triggers. *American Political Science Review 98*(1), 35–49.

Sohn, K. (2011). Acting White: A Critical Review. *Urban Review 43*(2), 217–234.

Storesletten, K. (2000). Sustaining Fiscal Policy through Immigration. *Journal of Political Economy 108*(2), 300–323.

Tyson, K. (2006). *Beyond Acting White: Reframing the Debate on Black Student Achievement*, Chapter The Making of a "Burden": Tracing the Development of a "Burden of Acting White" in Schools. Eds. E. Hovart and C. O'Connor. Rowman and Littlefield.