Access Econ
Scholarly communication without the middle men

Submission Number: PET11-11-00026

A theory of discrimination and assimilation

Jon X Eguia
New York University

Abstract

In a heterogeneous society with two social groups possessing competing social norms, members of the relatively worse-off group face an incentive to adopt the social norms of the better-off group and assimilate into it. I present a theory in which the cost of assimilation is endogenous and strategically chosen by the better-off group in order to screen those who wish to assimilate. In equilibrium, only high types, who generate positive externalities to the members of the better-off group, assimilate. As an application of the theory, I show that the so called "acting white" phenomenon in which students of a disadvantaged ethnic group punish peers who succeed academically can be explained as an optimal strategy on the part of untalented students to try to keep their more able co-ethnics in their community.

Submitted: February 10, 2011.

A Theory of Discrimination and Assimilation*

Jon X Eguia[†]

New York University

February 8, 2011

Abstract

In a heterogeneous society with two social groups possessing competing social norms, members of the relatively worse-off group face an incentive to adopt the social norms of the better-off group and assimilate into it. I present a theory in which the cost of assimilation is endogenous and strategically chosen by the better-off group in order to screen those who wish to assimilate. In equilibrium, only high types, who generate positive externalities to the members of the better-off group, assimilate. As an application of the theory, I show that the so called "acting white" phenomenon in which students of a disadvantaged ethnic group punish peers who succeed academically can be explained as an optimal strategy on the part of untalented students to try to keep their more able co-ethnics in their community.

Keywords: Assimilation, discrimination, social groups, "acting white".

*I am indebted to Will Terry for his extensive comments and advice on this project. I thank Renee Bowen, Autumn Carter, Rachel Kranton, Dimitri Landa, Ryan Pevnick, Carlo Prato, Jakub Steiner, Leonard Wantchekon and attendants at talks at LSE and Kellogg for their suggestions, and the Ford Center for Global Citizenship and the Center for Mathematical Studies at the Kellogg School of Management (Northwestern U.) 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 an unequal society with deep cultural or ethnic cleavages, in which one social group is privileged and others are not, members of disadvantaged groups face an incentive to embrace the dominant culture and assimilate into the most advantaged social group. I address two intimately related questions: When is it optimal for minorities to assimilate, given that assimilating is a costly enterprise? If members of the advantaged group are purely selfish, how receptive or hostile are they toward assimilation?

I present a theory of social integration in a society comprised of two groups: An advantaged group exogenously endowed with favorable status or wealth, and a disadvantaged group that lacks this wealth and status. Agents are characterized by their group of origin, and their ability. Agents generate externalities for members of the group to which they ultimately belong; wealthier and more able agents generate more positive externalities. Disadvantaged agents choose whether or not to they will assimilate by joining the advantaged group. Advantaged agents choose how difficult it is to assimilate and join their group.

I find that advantaged agents optimally screen those who seek to assimilate by choosing a difficulty of assimilation such that the agents who assimilate are precisely those whose ability is 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 gap in wealth or status between groups. The theory rationalizes using some cultural norms such as rules about language use and rejecting other individual traits such as skin color or place of birth to determine membership into the most advantaged social group: norms and traits that highly able individuals can acquire at a lower cost than less able individuals make it possible to screen those who seek to assimilate according to ability, so that only the most able individuals assimilate, which is the optimal outcome for the advantaged group.

An application of the theory serves to explain 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, black and Latino students who obtain top grades are *less* popular than black and Latino students with lower grades (Fryer and Torelli [22]). The traditional explanation (Fordham and Ogbu [17] and Fordham [16]) is cultural: African-Americans embrace academic failure as part of their identity and shun they those who defy this identity by studying, and the rationale for this defeatist identity is that society denied African-Americans career opportunities and did not reward their effort. McWhorter [34] 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.

There are two problems with these accounts. First, the premise that African-Americans see no reward for their efforts may have been true in the past, but it is false today: Education pays off for African-Americans,[1] and African-Americans know that effort pays off.[2] Second, as Fryer [18] points out, the identity and self-sabotage explanations imply that the acting white effect ought to be more prevalent in more segregated schools and in schools with fewer inter-ethnic friendships, where the black identity is stronger. Alas, the opposite holds true: The acting white problem is more acute among students who have greater exposure to whites (Fryer and Torelli [22]). Austen-Smith and Fryer [2] propose a micro-economic explanation based on the opportunity cost of studying. Students who are socially inept do not enjoy their leisure time, so they choose to study. Other

[1] As I discuss below, according to the US Census of 2000, a high school diploma increases the average earning an African-American by 57%, a college degree by 240%, and a professional degree by 532%.

[2] According to the Pew Center [7], since around 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.

students differentiate themselves from the socially inept by choosing not to study. While this argument is compelling, it applies to students of all ethnicities, and thus it does not capture the difference in the attitudes of black and white students at the heart of the acting white problem, which remains unexplained.

I develop a theory that explains why black and Latino students, but not whites, experience a negative correlation between popularity and top 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 talented individuals generate positive externalities for the group to which they ultimately belong, and since society makes assimilation too difficult for the less talented disadvantaged students, the second best outcome for these unfortunate students is to retain the more able co-ethnics in their community. They achieve this by punishing academic excellence in order to deter the brighter students from acquiring the skills that are necessary to assimilate. In particular, African-American overachievers who "act white" in high school, can later assimilate and for all practical purposes become members in good standing of mainstream white society. 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 brighter co-ethnics for *acting* white because acting white is a prologue to *becoming* white.

This explanation is consistent with the empirical findings that the acting white phenomenon is more prevalent in less segregated schools, where blacks have greater opportunities to interact with whites and thus to assimilate into white society. In these schools, where the risk of losing the brightest members of the underprivileged group is greatest, their less talented peers exert greater effort to deter exit from the group.

Beyond the specific case of explaining the acting white phenomenon, the broader theory is most applicable to social settings where an outsider, such as an immigrant, a member of an ethnic minority or a migrant, may assimilate and join mainstream society. Upon arrival to a large multi-ethnic city, 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 the city and in the country at large. 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 (King [31]) face a qualitatively similar choice.[3]

Assimilation requires the agent to acquire and embrace new social norms, habits and customs, which is costly, but it provides greater opportunities, and the cost of assimilation depends crucially on the attitude of the advantaged group that the migrant or immigrant seeks to join. Regarding the attitude of the advantaged social group, three recent empirical articles study the attitudes of Dutch and US citizens toward immigration. Hopkins [29] identifies conditions that make a community more likely to be hostile to immigration. Sniderman, Hagendoorn and Prior [40] find that Dutch citizens favor immigration by highly educated workers, and not by those who are only suited for unskilled jobs.

Hainmueller and Hiscox [23] 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

[3] Some populations were subject to coerced immigration and had no opportunity to integrate for many generations (i.e. the African-American community). It is only recently that their members can choose whether or not to assimilate into mainstream society.

and are opposed to low-skilled immigration. They review different economic theories of attitudes toward immigrations and they conclude that none explains their findings: "economic self-interest, at least as currently theorized, does not explain voter attitudes toward immigration." I present a theory that is fully consistent with these results: Economic self-interest leads low-skilled and high-skilled citizens alike to welcome only high skilled immigrants. While immigration leads to discrimination and social tensions as the native community seeks to deter many immigrants from assimilating, the successful integration of the brightest immigrants ultimately results in a creative and intellectual boom for the community (Putnam [37]).

The theory in this paper builds upon an extensive literature on theories of social identity formation, and empirical and theoretical work on interethnic relationships. For an interdisciplinary perspective on identity, see Hogg and Terry [27] and Hogg [26] (social psychology), the survey by Hill [24] (law and economics), and Jenkins's [30] overview of identity theories in the social sciences. Of particular relevance is the literature on identity economics (Akerlof and Kranton [1]; Bénabou and Tirole [5]; Bisin and Verdier [3] and [4] among others), which explains the adoption by poor minorities of strategies that punish productivity and achievement and/or preserve the ethnic customs of the minority. Minority agents embrace and pass on to their descendents identities that are anti-achievement (Akerlof and Kranton [1]), traditional (Bénabou and Tirole [5]) or ethnic (Bisin and Verdier [3]) because if they shed this identity and embrace the productive/modern/majority identity, they suffer an exogenously given cost.

The first question that I address in this paper is whether or not it is optimal for disadvantaged agents to assimilate. Identity economics theories teach us that given a sufficiently high exogenous cost of assimilation, it is not. These theories do not ask the second question that motivates my research: Why do advantaged agents discriminate against those who seek to assimilate? I propose a theory that recognizes that the difficulty of assimilation is endogenous: it depends on the actions of the agents in the advantaged group, who choose their actions optimally to suit their own selfish interests.

Shayo [38] presents a very general framework in which the utility of an agent depends on her individual payoff, the status of the group she identifies with, and the distance in traits from the individual to the average member of the group she identifies with. This distance depends on the actions of the agents. While Shayo [38] does not solve the general model, his framework has proved useful in applications to redistributive policies (Klor and Shayo [32]) and institutional design in an ethnically divided society (Penn [36]). My theory departs from Shayo [38] in at least two respects: While Shayo's agents are altruistic toward their own group, I study agents who are purely selfish in the tradition of standard rational choice. Second, Shayo uses an introspective notion of identity, an individual's concept of self, which may not coincide with other agents' view of the individual. Consider instead an external concept of identity: Regardless of what the agent thinks of herself, how does the individual act in society and what do other agents think of her as a result? This external identity determines the opportunities for friendship and social connections, and the externalities experienced by the agent, and is the focus of my study.

The central object of analysis in this approach is not so much the identity of individuals, i.e. their notion of self, but rather, their outward behavior. Instead of trying to discern whether members of ethnic groups think of themselves as fundamentally distinct from those of other ethnicities, researchers in this tradition want to know if agents learn a language to communicate (Lazear [33]), form friendships (Currarini, Jackson and Pin [9] and [10]; Fong and Isajiw [15]; Echenique, Fryer and Kaufman [12]; Patacchini and Zenou [35]; Marti and Zenou [11]), go on dates (Fisman, Iyengar, Kamenica and Simonson [14]) and marry (Fryer [21]) across ethnicities and races.[4] The focus is on behavior and interactions with others, not on an introspective concept of self.

The closest reference to the current paper is the Fryer's [20] theory of endogenous group choice. Agents face a repeated choice between a narrow group, which corresponds to the *disadvantaged* group in my theory, and a wider society, which corresponds to my *advantaged* group. The agent can choose to invest in skills that are group specific, or in skills that are valued by society at large. Agents who invest in group specific skills are akin to those who do not assimilate in my theory. Accumulation of group specific skills signals an expectation of renewed membership in the group at a later stage. Members of the group reward this expectation of future membership by greater cooperation with the agent. The theory in this paper and Fryer's [22] reach a common conclusion: Disadvantaged agents suffer pressure from their peers to acquire a lower level of human capital that is valued by society at large.

The crucial difference between the models is that Fryer's [22] is an infinitely repeated game, and he describes one equilibrium out of the many that exist under standard folk theorem arguments. Whereas, I capture the relevant insight in a simpler, static model that has a unique equilibrium. This model generates empirical implications that are consistent with the previously unexplained findings by Hainmueller and Hiscox [23] on attitudes toward immigration, and Fryer and Torelli [22] on the acting white phenomenon.

The rest of the paper is organized as follows.

In the first part of the paper, I present the results for the benchmark model. Informally, Proposition 1 says that the advantage agents set an intermediate level of discrimination that optimally screens disadvantaged agents, so that only the most able ones assimilate. After detailing the com-

[4] Friendships, dates and marriages are all positive interactions. I study societies where the alternatives are assimilation and peaceful seggregation. Societies where a more plausible alternative to assimilation is inter-ethnic conflict face a different strategic environment, discussed by Calvert [6] and Fearon and Laitin [13].

parative statics and welfare analysis of this result, I discuss various generalizations of the theory to account for preferences over race, and heterogeneity of preferences among agents.

In the second part of the paper, I apply the theory to explain the behavior of high school students who punish acting white in schools with mixed ethnicities. Proposition 4 implies that disadvantaged blacks and Latinos oppose the assimilation of their co-ethnics into the advantaged white community because talented individuals generate positive externalities for the group to which they ultimately belong, and in consequence they punish other blacks and Latinos for engaging in conducts such as obtaining good grades that lead to assimilation. In light of this finding, I propose a policy intervention to align the incentives of disadvantaged students with the academic success of their co-ethnics, which should eradicate the acting white problem and attenuate the achievement gap across ethnic groups in primary and secondary education.

Theory

Consider a society divided into two sets: A set $\mathcal{A}$ of advantaged agents, and a set $\mathcal{D}$ of disadvantaged agents. Each advantaged agent $i \in \mathcal{A}$ is endowed with wealth in quantity $w_{\mathcal{A}} > 0$, while each disadvantaged agent $i \in \mathcal{D}$ is endowed with wealth $w_{\mathcal{D}} = 0$. I interpret wealth very broadly, to include not only monetary or financial wealth, but also less tangible endowments such as status or local knowledge.

Let θ_i denote the type of agent i, and assume that types are private information. Assume sets $\mathcal{A}$ and $\mathcal{D}$ each contains a continuum of individuals, with types uniformly distributed in $[0, 1]$ in each set. Let a_i be the ability of agent i, and in the benchmark model, assume that $a_i = \theta_i$. I endogenize ability in an application in the next section.

Assume 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 behavior in social situations, eat certain foods and not others, and they spend their leisure times in specific activities. I assume that every advantaged agent 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 advantaged agents grow up embracing the advantaged norms as their own, whereas, disadvantaged agents are brought up according to the disadvantaged social norms. For instance, Anglo-white US citizens speak English, wear Western clothes and follow (American) football, baseball and basketball, while other ethnic groups in the country grow up exposed to alternative languages, dress, or sports.

I assume that while many disadvantaged agents have no alternative but to belong to the disadvantaged social group D, a fraction $\lambda > 0$ of disadvantaged agents can choose whether or not to join the advantaged social group A instead, by embracing its social norms. Let $\mathcal{D}_E \subset \mathcal{D}$ be this set of disadvantaged agents who choose their social group strategically. I assume that the distribution of types in $\mathcal{D}_E$ is the same as in $\mathcal{D}$ or $\mathcal{A}$ that is, uniform on $[0, 1]$. We can interpret this set as the set of teenagers or young adults who choose a lifestyle as they find their place in the world, in contrast to elder people who have already chosen a way of life, which is fixed and very difficult to change. This paper is concerned with the strategic choice that these agents face between joining social group A or joining social group D.

Note that I use calligraphic letters $\mathcal{A}$ and $\mathcal{D}$ to refer to the two sets in the original partition of society into advantaged and disadvantaged agents according to the exogenously given endowment of wealth, while the standard letters A and D denote the partially endogenous social groups that result from the endogenous group decisions by agents in $\mathcal{D}_E$.

Any disadvantaged agent $i \in \mathcal{D}_E$ can choose to belong to D at no cost, or she can learn to act according to the norms of the group A and then proceed to join A instead of D, but this learning is costly. Let $e_i \in \{0,1\}$ be the choice of agent $i \in \mathcal{D}_E$, where $e_i = 0$ denotes that i stays with the disadvantaged group D, and $e_i = 1$ denotes that agent i chooses to enter the advantaged social group A. If agent i chooses $e_i = 1$, I say that she *assimilates*. Let e denote the decisions to assimilate by all agents in $\mathcal{D}_E$. Formally, $e : [0,1] \to \{0,1\}$ is a mapping from type to assimilation decision. Given e, the composition of the social groups is $A = \mathcal{A} \cup \{i \in \mathcal{D}_E : e_i = 1\}$ and $D = \mathcal{D} \backslash \{i \in \mathcal{D}_E : e_i = 1\}$.

The cost of assimilating is $e_i d c(a_i)$, where e_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 $c(a_i)$ is a continuously differentiable, strictly positive, strictly decreasing function of the ability a_i, which captures the intuition that more able agents can learn to adapt at a lower cost.

I endogenize the value of d as follows. 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 essential cultural prerequisites of A, then d is high. I assume that a representative agent in $\mathcal{A}$ chooses d strategically, according to the incentives that I detail momentarily.

Agents derive utility from their wealth, from their ability, and from the externalities generated by the wealth and ability of other agents in their social group. Let

$$U_i(w_i, a_i, d, e) \equiv \psi(w_i, a_i) + u_i(d, e),$$

where U_i is the utility function of agent i, $\psi(w_i, a_i)$ is the direct utility that agent i obtains from her own ability and wealth, and $u_i(d, e)$ is the utility that agent i obtains as a result of the discrimination and assimilation decisions made by herself and other agents. Since wealth and ability are exogenous in the benchmark model, $\psi(w_i, a_i)$ can be ignored and the optimization problem of agent i is $\max_{\{e_i\}} u_i(d, e)$.

I assume that agents derive utility from the average wealth and ability of the agents in their group. Agents do not have others-regarding preferences, but there are externalities 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. Wealthier and more able agents generate more positive externalities to their friends and members of their group.

Formally, let w_A be the average wealth of agents in A. Note that $w_A \in [\frac{w_{\mathcal{A}}}{1+\lambda}, w_{\mathcal{A}}]$, where the lower bound is achieved if every $i \in \mathcal{D}_E$ assimilates, and the upper bound is achieved if none assimilate. The average wealth of agents in D is in any case 0. For any $J \in \{\mathcal{A}, \mathcal{D}, A, D\}$, let a_J be the average ability of agents in J. Note that if no agent assimilates, $a_A = a_D = a_{\mathcal{A}} = a_{\mathcal{D}} = \frac{1}{2}$.

Let $v(w, a)$ be the utility that an agent derives from the externalities coming from other agents in her group when the average wealth and ability of these agents are w and a. Then, any $i \in \mathcal{A}$, who by assumption belongs to A at no cost, receives utility from externalities $u_i(d, e) = v(w_A, a_A)$, whereas a disadvantaged agent $i \in \mathcal{D}_E$ attains:

$$u_i(d, e) = (1 - e_i)v(0, a_D) + e_i[v(w_A, a_A) - dc(a_i)]. \quad (1)$$

If $e_i = 0$, agent i does not assimilate and the utility from the externalities is just that of an agent in D, that is, $v(0, a_D)$; whereas, if $e_i = 1$, agent i assimilates and attains $u_i(d,e) = v(w_A, a_A) - dc(a_i)$. I assume that v is twice continuously differentiable and strictly increasing in both arguments. For $x, y \in \{w, a\}$, let v_{xy} denote the cross-partial derivative with respect to x and y. I assume that $v_{ww} < 0$, $v_{aa} \leq 0$ and $v_{wa} \geq 0$; the marginal utility of each variable is decreasing, and as the quantity of one variable increases, the other variable becomes more important ($v_{wa} \geq 0$ can be weakened, as long as the marginal utility of ability decreases less than the marginal utility of wealth when wealth increases).

I model the interaction of the agents as a game. The set of players is the set $\mathcal{D}_E$ and a single agent representative of the set $\mathcal{A}$ who seeks to maximize $v(w_A, a_A)$. The representative agent chooses the difficulty of assimilation level $d \in \mathbb{R}_+$, so her strategy space is $\mathbb{R}_+$. Each $i \in \mathcal{D}_E$ chooses whether or not to assimilate, so her strategy set is $\{0, 1\}$. The chosen strategies determine the average ability and wealth of each social group, and hence payoffs.

The timing is simple: First the representative advantaged agent selects $d \in \mathbb{R}_+$, which becomes common knowledge, then each $i \in \mathcal{D}_E$ chooses $e_i \in \{0, 1\}$, which determines the payoffs to every agent.

Results

I solve by backward induction.

Given d, and given any strategy profile by all other members of $\mathcal{D}_E$, an agent $i \in \mathcal{D}_E$ prefers to assimilate only if her ability a_i is high enough so that her cost of assimilating $c(a_i)$ is sufficiently small. It follows that for any d, there is a cutoff $a(d)$ in the level of ability such that members of $\mathcal{D}_E$ choose to assimilate if and only if their ability is above $a(d)$. An agent with ability equal to the cutoff is indifferent; I arbitrarily assume that the indifferent agent does not assimilate.[5]

[5] Since any ability level has zero mass in the population, the decision of the agent with ability equal to the cutoff

For any ability $a \in (0,1)$, let $d(a)$ be the degree of difficulty of assimilation that the representative advantaged agent must choose in order to have a become this cutoff, so that only agents with ability above a choose to assimilate.

I show that $d(a)$ is a function, not a correspondence, and I find sufficient conditions so that it is strictly increasing: If λ is small, few agents assimilate regardless of the cutoff, and thus the change in average wealth w_A and ability a_A as a result of a change in cutoff a is small. It follows that the individual incentives to assimilate depend mostly on the wealth gap $w_{\mathcal{A}}$ and the ability level a_i, and not on the the joint decisions of other agents. Agents with higher ability are willing to overcome a higher difficulty d.

Given $\frac{\partial d(a)}{\partial a} > 0$, the representative advantaged agent can choose d^* to maximize her utility by setting $d^* = d(a^*)$ such that

$$a^* = \arg\max_{\{a\}} v(w_A, a_A) \text{ s.t. } a_A(a) = \frac{1+\lambda-\lambda a^2}{2+2\lambda(1-a)} \text{ and } w_A(a) = \frac{w_{\mathcal{A}}}{1+\lambda(1-a)},$$

where w_A and a_A are the average wealth and ability of the agents in A as a function of a given that agents in $\mathcal{D}_E$ assimilate if and only if their ability is above a.

Proposition 1 *There exist $w' > 0$, $\lambda' > 0$, $d^* > 0$ and $a^* \in \left(\frac{1}{2}, 1\right)$ such that for any $w_{\mathcal{A}} < w'$ and $\lambda < \lambda'$, there exists a unique subgame perfect equilibrium, in which the endogenous difficulty of assimilation is d^* and disadvantaged agents in $\mathcal{D}_E$ choose to assimilate if and only if their ability is above a^*.*

This and all other proof are in the Appendix. Advantaged agents, through their representative agent, choose a level of difficulty of assimilation d^*. Given this level of assimilation, only disadvan-

is irrelevant.

taged agents in $\{i \in \mathcal{D}_E : a_i > a^*\}$ assimilate. Although these agents who assimilate bring down the average wealth of the advantaged social group, their high ability increases the average ability of the group, and for sufficiently high ability types, the positive effect through the increase in ability dominates, so the net effect of the assimilation of these high-ability disadvantaged agents is positive for advantaged agents.

If the wealth difference between sets $\mathcal{A}$ and $\mathcal{D}$ is not too large, at least the most able $i \in \mathcal{D}$ brings a positive externality to group A if she assimilates, hence the optimal difficulty of assimilation is intermediate: positive, to deter low ability agents from assimilating, but not too high, to encourage high ability agents to assimilate. A small λ guarantees that the argument holds for general classes of the cost and benefit functions $c(a)$ and $v(w, a)$: If λ is large, for some functional forms, there is a cascade effect by which if higher types assimilate, lower types assimilate as well, and group $\mathcal{A}$ may prefer to prevents this cascade by making assimilation prohibitively costly for all types.[6]

Discrimination by means of imposing a cost of assimilation $d^* > 0$ is a screening device that the advantaged agents use to separate low ability agents, who are not welcome to assimilate, from high ability agents, who are welcome to assimilate.[7]

The model is a variation on the seminal signaling model by Spence [41]: Members of the set $\mathcal{D}_E$ choose whether to invest in assimilation techniques. This investment is less costly to more able agents, so in equilibrium the agents separate: Highly able agents invest, and less able agents do not. Observing the investments, the advantaged agents accept into their social group those who invested at least d^*. In the original model by Spence [41], agents invest in education, which serves no other purpose but signaling high type, they are hired by a firm and there are no externalities.

[6] We can also guarantee that the result holds for any given λ by letting the cost function $c(a)$ be sufficiently convex.

[7] By imposing a cost of assimilation, advantaged agents both *discriminate against* all disadvantaged agents, and –in a more favorable sense of the word- they *discriminate among* disadvantaged agents, by discerning and distinguishing who are the high types among them, who then proceed to assimilate.

In my model, agents invest in learning the patterns of behavior consistent with membership in a social group that it is not initially their own.

To study the comparative statics with changes in the wealth gap between groups, I relax the normalization that $w_{\mathcal{D}} = 0$, assuming instead that $0 \leq w_{\mathcal{D}} \leq w_{\mathcal{A}}$, so that I can study the effect of increases in the wealth of each group independently. Two preliminary results hold for extreme cases: If $w_{\mathcal{A}} = w_{\mathcal{D}}$, then there is no incentive to assimilate even if $d = 0$ because both groups are identical, so there is no strategic reason to set a positive d. If $w_{\mathcal{A}} - w_{\mathcal{D}}$ is too large, then d^* must be high enough to deter even the most able disadvantaged agent from assimilating. For a positive but not too great wealth gap, the wealth of the disadvantaged group is the more relevant factor: Even if the wealth 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. On the other hand, comparative statics on the effect of changes in the wealth of the advantaged group are indeterminate.

Let $a^*(w_{\mathcal{A}}, w_{\mathcal{D}})$ and $d^*(w_{\mathcal{A}}, w_{\mathcal{D}})$ be the equilibrium ability cutoff and difficulty of assimilation as a function of the wealth of each group, and let $\vartheta^*(w_{\mathcal{A}}, \Delta)$ and $\zeta^*(w_{\mathcal{A}}, \Delta)$ be the equilibrium cutoff types and difficulty of assimilation as a function of wealth, keeping the wealth gap $\Delta = w_{\mathcal{A}} - w_{\mathcal{D}}$ constant. The first pair of functions makes it possible to study comparative statics with respect to changes in wealth idiosyncratic to each group, whereas the second is useful to study the effect on increases of wealth across groups that do not change the wealth gap.

Proposition 2 *There exist $\Delta' > 0$ and $\lambda' > 0$ such that for any wealth gap $w_{\mathcal{A}} - w_{\mathcal{D}} < \Delta'$ and any $\lambda < \lambda'$,*

i) $a^*(w_{\mathcal{A}}, w_{\mathcal{D}})$ and $d^*(w_{\mathcal{A}}, w_{\mathcal{D}})$ are strictly decreasing in $w_{\mathcal{D}}$ for any $w_{\mathcal{D}} \in [w_{\mathcal{A}} - \Delta', w_{\mathcal{A}}]$, and

ii) $\vartheta^*(w_{\mathcal{A}}, \Delta)$ and $\zeta^*(w_{\mathcal{A}}, \Delta)$ are strictly decreasing in $w_{\mathcal{A}}$ for any $w_{\mathcal{A}}$ and any $\Delta \in [0, \Delta']$.

Furthermore, if $\lim_{w->\infty} \frac{\partial v(w,a)}{\partial w} = 0$*, then for any* $\Delta > 0$*,* $\vartheta^*(w_{\mathcal{A}}, \Delta)$ *and* $\zeta^*(w_{\mathcal{A}}, \Delta)$ *are non increasing in* $w_{\mathcal{A}}$ *for any* $w_{\mathcal{A}}$*, and there exists* w'' *such that* $\vartheta^*(w_{\mathcal{A}}, \Delta)$ *and* $\zeta^*(w_{\mathcal{A}}, \Delta)$ *are strictly decreasing for any* $w_{\mathcal{A}} > w''$*.*

The first result says that if the wealth gap is not too large, assimilation increases with the levels of wealth of the disadvantaged agents. The second and third result note that assimilation also increases if both groups become richer, keeping the wealth gap constant. Notice however that as the wealth of both groups increases, the inequality in the income distribution as measured by the Gini coefficient, or any other measure of relative wealth is reduced, so this result does not imply that assimilation occurs in a society that becomes richer while keeping its levels of relative inequality constant.

Welfare analysis with respect to the difficulty of assimilation d is not straightforward. Advantaged and disadvantaged agents have conflicting interests: Advantaged agents want disadvantaged agents with high ability to assimilate, but this assimilation makes the other disadvantaged agents worse off. The utility of advantaged agents achieves a global minimum at 0 where every $i \in \mathcal{D}_E$ assimilates, and it is first strictly increasing up to the equilibrium value d^*, then strictly decreasing up to the level $d(1)$ where nobody assimilates. The utility of agents who stay in the disadvantaged group is first strictly decreasing and then strictly increasing up to $d(1)$. The utilities of those who assimilate diverge according to ability, because the cost they incur depends on their ability.

Aggregate welfare analysis depends on interpersonal comparisons, and on the weight assigned to each group. In equilibrium, and compared to the benchmark with no assimilation, advantaged agents and the agents with the highest types among those assimilate benefit from assimilation, while disadvantaged agents who do not assimilate and those with the lowest types among those who assimilate become worse off. A level $d = 0$ uniquely maximizes the utility of disadvantaged agents (including those who assimilate), and it minimizes inequality, but it minimizes the utility of advantaged agents.

In the next subsection I consider four generalizations: (1) distinguishing between costs of assimilation based on norms and costs based on exogenous traits such as race; (2) allowing for preferences over these exogenous attributes; (3) discussing a more symmetric model in which $\mathcal{A}$ and $\mathcal{D}$ are each endowed with a different kind of wealth, so that assimilation and discrimination occur in both directions; and (4) generalizing the utility functions so that the externalities that an agent receives depend on her ability, which generates a conflict of interest between low ability and high ability agents in $\mathcal{A}$.

Generalizations

One – In a society where the division of wealth corresponds to an ethnic divide of the population, 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 of wealth. 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 the advantaged agents seek to harness the positive externalities provided by highly able individuals, an optimal discrimination policy should 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 ability.

Put it differently, even if advantaged agents care only about their self-interest and are unconcerned about the welfare of disadvantaged agents, discrimination based on exogenous attributes is not in the best interest of advantaged agents because exogenous attributes do not provide information about type.

To formally this argument, let d_R be an additional cost that all those who assimilate must pay, and let a representative $i \in \mathcal{A}$ choose d and d_R. The cost d_R is based on the exogenous attributes of the individual, such as race or place of birth: It is a lump sum cost.[8] Then for any $i \in \mathcal{D}_E$, instead of expression 1, we now have:

$$u_i(d,e) = (1-e_i)v(0,a_D) + e_i[v(w_A,a_A) - dc(a_i) - d_R]. \tag{2}$$

As long as d is positive, screening works, and there is a substitution effect between imposing costs by means of d or d_R. However, the screening effect by which low types are deterred from assimilation is weaker as d is substituted by d_R. If we introduce some noise into the model, advantaged agents strictly prefer stronger screening. Introduce an ξ_i noise independently drawn from a uniform distribution over $[-\xi, \xi]$, so that expression 2 becomes:

$$u_i(d,e) = (1-e_i)v(0,a_D) + e_i[v(w_A,a_A) - dc(a_i) - d_R + \xi_i]. \tag{3}$$

Recall a^* and d^* denote the equilibrium assimilation cutoff and difficulty in the benchmark model. In the generalization with d_R and noise, given d and d_R such that agent with type a^* is in expectation indifferent between assimilating or not, ex post some agents with higher ability do not

[8] Unlike d, the cost d_R is discriminatory only in the pejorative sense of the word. Advantaged agents *discriminate against* (and not among) disadvantaged agents by imposing the cost d_R.

assimilate while other agents with ability below a^* assimilate, reducing the utility of advantaged agents. The quality of the screening is increasing in d and invariant in d_R. Thus, in the extended game with utility function 3, advantaged agents achieve the best screening by choosing the lowest possible d_R and the highest possible d consistent with $d(a^*)+d_R = d^*$. Let d_R^{**} denote the equilibrium value of d_R in the game with $u_i(d, e)$ given by expression 3.

Assume that the set of feasible values of d_R is an interval $[d_R^-, d_R^+]$. In equilibrium, discrimination based on her exogenous traits is set to the minimum feasible value.

Claim 3 *Suppose $a^* < 1$. In the unique equilibrium to the game with $u_i(d, e)$ given by expression 3, $d_R^{**} = d_R^-$.*

Advantaged agents, if they are strategic, do not discriminate on the basis of immutable characteristics such on skin color, race, place of birth. Rather, the advantaged group prefers to screen on the basis of some observable characteristic that correlates with ability. Advantaged agents can construct and use a set of norms that are naturally less costly to acquire for disadvantaged high types, and then they can adopt a simple cut-off rule: The disadvantaged agents who acquire a sufficiently high proficiency in the set of norms of $\mathcal{A}$ must be a sufficiently high type, and thus she should be assimilated, while disadvantaged 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.

Two – 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 advantaged agents dislike some exogenous attribute of set $\mathcal{D}$, advantaged agents treat the disadvantaged as if the wealth differential was higher, the equilibrium difficulty of assimilation d^* rises, and fewer agents assimilate. If the disadvantaged agents dislike some exogenous attribute of $\mathcal{A}$, then the disadvantaged perceive the wealth difference as smaller, and the equilibrium difficulty of assimilation d^* must be lower in order to entice the disadvantaged 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 (Hong and Page [28]), in equilibrium there is less discrimination and more assimilation.[9]

Three – Alternatively, we can 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, w and m. Every $i \in \mathcal{A}$ is endowed with w in quantity $w_{\mathcal{A}}$ and every $i \in \mathcal{D}$ is endowed with m in quantity $m_{\mathcal{D}}$, while $w_{\mathcal{D}} = m_{\mathcal{A}} = 0$. As before, a set $\mathcal{D}_E \subset \mathcal{D}$ of size λ can assimilate into A. To make the model symmetric, let a set of agents $\mathcal{A}_E \subset \mathcal{A}$ choose whether to assimilate into D. The reason to assimilate into D is that within $\mathcal{A}_E$ and $\mathcal{D}_E$, a fraction γ of agents are not interested in wealth w and receive no utility from it. Instead, this fraction of agents value m and receive positive externalities from the average ability and average value of m in the social choice group they join. Representative agents from $\mathcal{A}$ and $\mathcal{D}$ each choose

[9] The formalization and proof of the results for generalizations two, three and four are available from the author.

a level of discrimination d_A and d_D and those who to assimilate into $J \in \{A, D\}$ must incur a cost $d_J c(a_i)$.

Every agent i who values wealth w 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 $i \in \mathcal{D}_E$, then i assimilates if and only if a_i is sufficiently high. However, now assimilation goes both ways: Agent $i \in \mathcal{A}_E$ who values m assimilates into D if and only if her ability is sufficiently high.

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

Four $-$ Another generalization of the model is to let w be the only wealth and assimilation to occur only from $\mathcal{D}_E$ to A, but to introduce heterogeneity in the preferences of agents in $\mathcal{A}$ over assimilation by agents in $\mathcal{D}_E$. Suppose that the externality of a high average ability in a social group is greater for more able agents. The intuition is that it takes intelligence to enjoy the intelligence of others. The formalization is that $v_i(w_A, a_A)$ must depend on a_i so that $v_i(w_A, a_A) = v(w_A, a_A, a_i)$, where $v_{wa_i} = 0$ and $v_{a_A a_i} > 0$. Under this generalization, all advantaged agents agree on the disutility of letting poor agents assimilate, but highly able agents are keener to let other highly able agents assimilate. As a result, advantaged agents disagree on the optimal level of discrimination and the optimal level of assimilation: Highly able agents want less discrimination and more assimilation to increase the average ability level in social group A. Less able agents want more discrimination and less assimilation to preserve the high wealth level in group A.

Application: Acting White

Fryer [19] defines "acting white" as "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." He 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 black and Hispanic students relative to white students.

I argue that punishing high achieving disadvantaged students is a fully rational, best response strategy by less talented disadvantaged students.

To explain the "acting white" phenomenon, I let the ability of an agent be partially endogenous, depending on both the type of the agent, and the agent's choice of education. Recall that θ_i is the type of agent i and the distribution of types is uniform $[0, 1]$ in both $\mathcal{A}$ and $\mathcal{D}$. Given her type, agent i chooses $x_i \in [0, \theta_i]$, and the resulting ability a_i is realized by Nature by a random draw from a uniform distribution in $[(1 - \varepsilon)x_i, x_i]$ for some arbitrarily small ε.[10]

I interpret type θ_i as the potential or talent which serves as an upper bound on the skills and ability that agent i can acquire. I interpret x_i as the choice to acquire a level of skills or education, and this choice x_i, together with a small noise from Nature, determines the ability a_i of the agent. Individuals who waste away their school years do not achieve their potential, they come out of school with fewer skills, and are less able to succeed in society. They have become, by choice, identical to agents with a lower type.

Recall the utility of agent i is $\psi(w_i, a_i) + u_i(d, e)$. All else equal, every i prefers the highest possible level of skills x_i to maximize a_i and $\psi(w_i, a_i)$. But all else is not equal: In some schools,

[10] The stochastic nature of the ability is a technical assumption to guarantees that the distribution function of ability levels in the population is continuous. Note that $\varepsilon > 0$ is arbitrarily small, so that a_i is arbitrarily close to x_i.

peers may punish those who excel.

I introduce peer pressure into the theory. As described in the main section, a set $\mathcal{D}_E$ comprising a fraction λ of disadvantaged agents choose strategically the social group they want to belong to. I interpret the set $\mathcal{D}_E$ as the set of young disadvantaged agents with some contact with advantaged agents so that they have an opportunity to imitate the behavior of the advantaged group, internalize its norms and assimilate. Assume that disadvantaged agents who belong to $\mathcal{D}_E$ are susceptible to peer pressure by their peers. For symmetry, assume as well that a set $\mathcal{A}_E \subset \mathcal{A}$ of size λ of advantaged agents are susceptible to peer pressure by other advantaged agents.

Peer pressure takes the following form: An agent $j \in \mathcal{D}$ with low type $\theta_j < \frac{1}{2}$ chooses a threshold $x_{\mathcal{D}}^P \in [0, 1]$ and every $k \in \mathcal{D}_E$ who chooses $x_k > x_{\mathcal{D}}^P$ is punished and has to pay a fixed cost $K > 0$. Similarly, a representative low type $i \in \mathcal{A}$ chooses a threshold $x_{\mathcal{A}}^P$ to punish any $k \in \mathcal{A}_E$ such that $x_k > x_{\mathcal{A}}^P$.[11] We can interpret these punishments as physical bullying, or more mildly, we can imagine that disadvantaged individuals are rationally more keen to invest costly resources developing friendships with individuals who are more likely to stay in the group, isolating those who are likely to leave the community.

The utility function of an agent $k \in \mathcal{D}_E$ is:

$$\begin{aligned} \psi(w_k, a_k) + (1 - e_k)v(0, a_D) + e_k[v(w_A, a_A) - dc(a_i)] - K \text{ if } x_k &> x_{\mathcal{D}}^P, \\ \psi(w_k, a_k) + (1 - e_k)v(0, a_D) + e_k[v(w_A, a_A) - dc(a_k)] \text{ if } x_k &\leq x_{\mathcal{D}}^P. \end{aligned} \tag{4}$$

I keep the same assumptions about functions v and c as in the benchmark model. The set of players in this new game is larger:

[11] All low types in their respective groups share the same preferences over $x_{\mathcal{D}}^P$ and $x_{\mathcal{A}}^P$, so we can let a representative agent of each group choose the threshold.

-The representative $i \in \mathcal{A}$ with $\theta_i < \frac{1}{2}$ who chooses d and $x_{\mathcal{A}}^P$.

-The representative $j \in \mathcal{D}$ with $\theta_j < \frac{1}{2}$ who chooses $x_{\mathcal{D}}^P$.

-The set of agents $\mathcal{D}_E$ who choose whether to assimilate and their skill level.

-Every other agent, who each chooses her own skill level.

The timing is as follows:

1. An agent $j \in \mathcal{D}$ with low type chooses the peer pressure threshold $x_{\mathcal{D}}^P$ and, simultaneously, an agent $i \in \mathcal{A}$ with low type chooses the peer pressure threshold $x_{\mathcal{A}}^P$ and the difficulty of assimilation d.[12]

2. Each agent k chooses her skill level $x_k \in [0, \theta_k]$, and nature determines her ability a_k, distributed uniformly in $[(1-\varepsilon)x_k, x_k]$. Any $k \in \mathcal{A}_E \cup \mathcal{D}_E$ who chooses x_k above the threshold for her group incurs punishment K.

3. Agents in $\mathcal{D}_E$ choose whether to assimilate or not, determining the payoffs for every agent.

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

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

At step 2, any agent $k \notin \mathcal{A}_E \cup \mathcal{D}_E$ chooses education $x_k = \theta_k$, their true potential. Any $k \in \mathcal{A}_E$ chooses $x_k \in \{\theta_k, x_{\mathcal{A}}^P\}$ and any $k \in \mathcal{D}_E$ chooses $x_k \in \{\theta_k, x_{\mathcal{D}}^P\}$.

At step 1, the representative advantaged low type has no incentive to punish any advantaged agent, because a higher level of ability of any $k \in \mathcal{A}$ generates positive externalities to all members of $\mathcal{A}$. Hence $x_{\mathcal{A}}^P = 1$ and no agent in $\mathcal{A}$ ever suffers peer pressure to become less able.

Whereas, agent $j \in \mathcal{D}$ who chooses $x_{\mathcal{D}}^P$ has an incentive to lower the ability of some agents to prevent them from assimilating. Let Θ be an arbitrary pair of continuous distributions of levels of

[12] The result is robust to variations in the timing of moves. For a model where d is chosen before $x_{\mathcal{D}}^P$ and $x_{\mathcal{A}}^P$, see working paper versions of the manuscript from 2010.

ability in $\mathcal{A}$ and $\mathcal{D}$. For any Θ, there is a threshold function increasing in d such that in equilibrium of the subgame that follows given (d, Θ) disadvantaged agents choose to assimilate if and only if their ability is above the threshold. In equilibrium, disadvantaged low types are hurt by this assimilation. Fixing $x_{\mathcal{D}}^P$ below the threshold of assimilation deters some agents in $\mathcal{D}_E$ from acquiring a level of ability above the threshold and thus from assimilating. The optimal peer pressure maximizes w_D by making as many disadvantaged high types stay in group D as possible, while lowering their ability only just as much as it is necessary to prevent them from assimilating, and no more.

Proposition 4 *There exist $w' > 0$ and $\lambda' > 0$ such that for any $w_{\mathcal{A}} < w'$ and $\lambda < \lambda'$, in all subgame perfect equilibria, $x_{\mathcal{D}}^P < 1 = x_{\mathcal{A}}^P$; that is, acting white occurs in equilibrium.*

High type disadvantaged agents are pressured to underperform and become less able, whereas no advantaged agent endures this pressure. This is the "acting white" phenomenon.

Note that the equilibrium d is lower than without peer pressure: Fewer high types assimilate, and as a consequence, the average ability in A is lower, so intergroup differences are smaller, making assimilation less desirable. I illustrate these and other differences with a numerical example, and a figure that captures the qualitative results.

Example 5 *Let $w_{\mathcal{A}} = 4$, $\lambda = 0.1$, $c(a_i) = \frac{1}{a_i}$, $\psi(w_i, a_i) = w_i^{1/2} + 10a_i$, $v(w, a) = w^{1/2} + 10a$. 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 compare the equilibrium outcomes under an assumption of not peer pressure ($K = 0$) in column 2, and peer pressure ($K = 1$) in column 3, where $x_{\mathcal{D}}^P = 0.6$ as part of the equilibrium.*

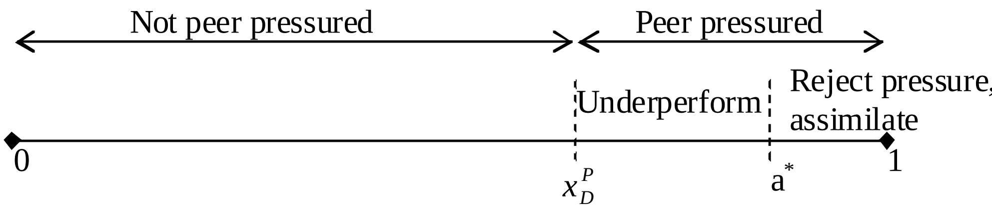

Figure 1: Incidence of Acting White phenomenon on disadvantaged agents, by type.

	No peer pressure	*Peer pressure*	*(3)-(2)*
d^*	*1.341*	*1.314*	*-0.027*
a^*	*0.610*	*0.676*	*0.066*
a_A	*0.511*	*0.510*	*-0.001*
a_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***

While the acting white equilibrium makes all advantaged agents worse off and it reduces the average utility of disadvantaged agents and aggregate welfare, it makes disadvantaged agents with low types –the perpetrators of peer punishments- better off.

Figure 1 summarizes the effects of the acting white phenomenon on the disadvantaged agents. The horizontal axis measures type or innate talent. The not so talented students with type below $x_{\mathcal{D}}^P$ are not subjected to any peer pressure. The talented students with type above $x_{\mathcal{D}}^P$ are subjected to peer pressure to underperform. Those with type between the punishment threshold $x_{\mathcal{D}}^P$ and the equilibrium assimilation cutoff a^* yield to the pressure and underperform to escape social punishments, while the most talented reject the peer pressure, endure the consequent alienation from their co-ethnics, and ultimately assimilate into the advantaged community. This prediction is consistent with and can explain the bimodality of the empirical distribution of black students by number of white friends in integrated schools: "There are, mainly, two types of black students: those who have mostly white friends and those who choose mostly black friends"(Patacchini and Zenou [35]; Marti and Zenou [11]). Those with white friends, I theorize, are those on their way to assimilate.

The "acting white" phenomenon is asymmetric: Disadvantaged minorities (typically identified as blacks students in the literature that focuses in the US) sabotage their peers in their acquisition of skills; whereas, mainstream social groups (typically categorized as "whites") do not sabotage their peers in their acquisition of skills. This asymmetry is missed in the explanation given by Austen-Smith and Fryer [2], according to which high-schoolers shun studious colleagues because studiousness signals a lack of social graces. While their argument is compelling, it applies equally to all races and social groups.

Fordham and Ogbu [17] and Fordham [16] argue that the asymmetry is part of what defines the identity of the groups: "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 through the accumulation of the skills taught at school. Fordham and Ogbu [17] find that students in the 1980s perceived activities such as speaking standard English, getting good grades, or going to libraries or museums, as distinctly "white" and they stress that to engage in these behaviors is to give up membership in the black social group, to "join the enemy." 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 acting white problem is most severe in integrated schools in which blacks have the greatest opportunities to climb in the social ladder (Fryer [19]); whereas, the oppositional culture theory would imply that the problem ought to be worse in schools with the least socioeconomic opportunities. 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 compared to the income of blacks who do not finish high school, the average monthly income for blacks is 57% higher for those with a high school diploma, 129% for an associate degree from college, 240% for a bachelor degree, 298% for a master degree, and 532% for a professional degree. Academic success pays off for today's black students, even if Fordham and Ogbu [17] are right that it formerly did not.

Even though black students accept that they are as capable of learning as whites, and society rewards students of any ethnic background for their academic achievements, black (and Latino) students continue to be punished by their peers for pursuing academic success in high school, while white students are not.

The theory in this paper explains this asymmetry while keeping symmetric assumptions on the utility function of the agents, the distribution of types and the technology for peer pressure. Solely from an initial inequality in the allocation of wealth, it logically follows that some poor agents choose to sabotage their peers's acquisition of skills, but rich agents do not. Such seemingly self-hurting behavior is in fact a fully rational strategy that aims to retain talent within the disadvantaged group.

The term "self-sabotage" (McWhorter [34]) improperly anthropomorphizes the African-American minority as a group that hurts itself. To the extent that self-saboteurs are deemed unworthy of social assistance, this description has important normative consequences. And yet its use is misleading at best, and arguably tendentious. No individual African-American engages in self-sabotage: When some African-Americans penalize other students' academic success, they are best responding to the incentive structure created by mainstream society.

The punishment of high achieving African-American students is only an instance of a broader social phenomenon: Hoff and Sen [25] report a strikingly phenomenon 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. Students in rural schools face an analogous strategic environment, since academic success leads to migration to the city; therefore, I conjecture that rural students who obtain top grades also lose popularity, regardless of their race. If true, this is akin to the "acting white" phenomenon, or, in this case, "acting urban."

Another implication is that 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.

As noted above, Fryer [19] and Fryer and Torelli [22] find that the acting white problem is more severe in less segregated schools: In predominantly black schools, "there is no evidence at all that getting good grades adversely affects students' popularity" (Fryer [19]). Fryer and Torelli [22] find this "surprising," but my theory offers and explanation for this result: Only black students in mixed schools are exposed to interaction with white students, so these students –as opposed to those in segregated schools- have a greater opportunity to have 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 [19] rhetorically conjectures that perhaps the problem is attenuated and solved if school desegregation leads to cross-ethnic friendships. My theory suggests the opposite: The greater the influence of white culture over black students, the greater the risk that the best black students may choose to assimilate into white society and leave behind the black students who are not as talented and would thus not be as well received by the dominant culture in society. Fryer [19] reports that indeed, greater inter-ethnic integration leads to a more severe acting white problem.

Policy recommendations

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 brightest classmates.

If the classmates of an honors student perceive it to be in their immediate interest that the top student excels, they will see to it that they do not punish success. Disdain for academic success is not circumscribed to minority students: Coleman [8] found in the 1950s that cheerleaders and athletes, not those with the best grades, were the most popular students. Athletes' efforts result in honor and glory for the whole school. Whereas, students toil for their own individual gain. There is very little spillover for her classmates and neighbors if a high-schooler from a marginalized neighborhood succeeds against all odds in high school, obtains a scholarship to go to college, and moves thousands miles to start a new life in an elite university.

Conditional cash transfers or other policies that provide rewards based on observed behavior or outcome can change individual incentives in the classroom setting. Slavin [39] details the results of an international survey of financial incentives schemes aimed to increase education achievements. He finds that these schemes have positive results in developing countries, but do not seem to work as well 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, or passes an exam, etc), without any attention to peer effects. These financial incentives therefore reinforce the perception of educational achievement as a purely individualistic good. It is possible that these incentives could in fact make the acting white phenomenon worse: If minority students lose popularity for studying, would they not lose even more popularity for studying for a small reward from education authorities?

I suggest instead that conditional financial incentives that are distributed to a group of peers, and not to an individual, may be more effective in mitigating the punishment of high achievement and reducing the achievement gap in education between ethnic groups. A program that rewards *every* student in a class with a cash transfer that is contingent on the *total* number of A grades obtained by the collective body of students in the class changes educational achievement from an individualistic good that betrays an aspiration to abandon the community, into a team production good that immediately benefits every member of the community, by means of the contingent collective financial incentive. I conjecture that under these incentives, the A students who produce the public good enjoyed by all their classmates would no longer lose popularity for achieving their high grades and delivering these public goods.

Appendix

Proposition 1

Proof. Let $a_A(a)$ and $w_A(a)$ be the average ability and wealth in A and let $a_D(a)$ be the average ability of agents in D as a function of a given that agents in $\mathcal{D}_E$ assimilate if and only if their type is above a. Then

$$\begin{aligned}
w_A(a) &= \frac{w_{\mathcal{A}}}{1+\lambda(1-a)}, \qquad (5)\\
a_A(a) &= \left[\frac{1}{2}+\lambda(1-a)\frac{1+a}{2}\right]\frac{1}{1+\lambda(1-a)} = \frac{1+\lambda-\lambda a^2}{2+2\lambda(1-a)},\\
a_D(a) &= \left[a\frac{a}{2}+(1-\lambda)(1-a)\frac{1+a}{2}\right]\frac{1}{a+(1-\lambda)(1-a)} = \frac{1-\lambda+\lambda a^2}{2-2\lambda(1-a)}.
\end{aligned}$$

Given any d and any strategy profile e_{-i} for every $j \in \mathcal{D}_E \backslash \{i\}$, since $c(a_i)$ is strictly decreasing in a_i, agent i chooses $e_i = 1$ if and only if a_i is above some cutoff that depends on d and e_{-i}. For any $i, j \in \mathcal{D}_E$ such that $a_i > a_j$, and given any d and any strategy profile $e_{-i,j}$ for every $h \in \mathcal{D}_E \backslash \{i, j\}$, if i and j best respond, $e_j = 1$ implies $e_i = 1$. Hence, given any d, there exists a unique cutoff in $[0, 1]$ such that for any $i \in \mathcal{D}_E$, $e_i = 1$ if and only if a_i is above the cutoff, which depends on d. Let $d(a)$ be the degree of difficulty that makes a be this cutoff. Let $v(w, a)|_{w=x,a=y}$ denote the value of $v(w, a)$ evaluated at $w = x$ and $a = y$. Then

$$d(a) = \frac{v(w,\vartheta)|_{w=w_A(a),\vartheta=a_A(a)} - v(w,\vartheta)|_{w=0,\vartheta=a_D(a)}}{c(a)}.$$

Note that if $\lambda = 0$, then

$$d(a) = \frac{v(w,\vartheta)|_{w=w_{\mathcal{A}},\vartheta=\frac{1}{2}} - v(w,\vartheta)|_{w=0,\vartheta=\frac{1}{2}}}{c(a)},$$

which is a strictly increasing, continuously differentiable function, with

$$d'(a) = -\frac{v(w,\vartheta)|_{w=w_{\mathcal{A}},\vartheta=\frac{1}{2}} - v(w,\vartheta)|_{w=0,\vartheta=\frac{1}{2}}}{[c(a)]^2}c'(a) > 0.$$

Note that for any $\lambda \in [0, 1)$, since $w_A(a)$, $a_A(a)$, $a_D(a)$, $c(a)$, $c'(a)$ are continuous in λ for any

$\lambda \in [0, 1)$, $v(w, \vartheta)$ is continuous, and $c(a)$ is positive for any a, so both $d(a)$ and $d'(a)$ are continuous in λ for any $\lambda \in [0, 1)$. Therefore, there exists $\lambda' > 0$ such that $d'(a) > 0$ for any $\lambda < \lambda'$. Let

$$\begin{aligned} a^* &= \arg\max_{a \in [0,1]} v(w, \vartheta) \text{ s.t.} \\ w &= w_A(a) = \frac{w_{\mathcal{A}}}{1 + \lambda(1-a)}, \\ \vartheta &= a_A(a) = \frac{1 + \lambda - \lambda a^2}{2 + 2\lambda(1-a)}. \end{aligned} \tag{6}$$

Since $v(w_A(a), a_A(a))$ is continuous in a, it achieves a maximum on the compact set $[0, 1]$, so a solution exists. I first show that for a sufficiently low $w_{\mathcal{A}}$, the solutions must be interior. First, $a = 0$ is not a solution, because $\frac{dv(w_A, a_A)}{da} > 0$ at $a = 0$. Second, $a = 1$ is not a solution for a low enough $w_{\mathcal{A}}$, because if $a = 1$, then

$$\begin{aligned} \frac{dv(w_A, a_A)}{da} &= \lambda w_{\mathcal{A}} \frac{\partial v(w_A, a_A)}{\partial w_A} + \frac{-2\lambda(1+\lambda) + \lambda}{2} \frac{\partial v(w_A, a_A)}{\partial a_A} \\ &= \lambda w_{\mathcal{A}} \frac{\partial v(w_A, a_A)}{\partial w_A} + \frac{-\lambda - \lambda^2}{2} \frac{\partial v(w_A, a_A)}{\partial a_A} \end{aligned}$$

which is negative if

$$w_{\mathcal{A}} < \frac{1 + \lambda}{2} \frac{\frac{\partial v(w_A, a_A)}{\partial a_A}}{\frac{\partial v(w_A, a_A)}{\partial w_A}}.$$

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

$$\frac{dv(w_A, a_A)}{da} = \frac{\partial w_A}{\partial a} \frac{\partial v(w_A, a_A)}{\partial w_A} + \frac{\partial a_A}{\partial a} \frac{\partial v(w_A, a_A)}{\partial a_A} = 0. \tag{7}$$

Note that

$$\frac{\partial w_A}{\partial a} = \frac{\lambda w_{\mathcal{A}}}{[1 + \lambda(1-a)]^2} \text{ and } \frac{\partial a_A}{\partial a} = \frac{-2\lambda a[1 + \lambda(1-a)] + \lambda(1 + \lambda - \lambda a^2)}{2[1 + \lambda(1-a)]^2},$$

so a solution $a = a^*$ satisfies

$$\begin{aligned} 0 &= \frac{1}{[1+\lambda(1-a)]^2}\left(\lambda w_{\mathcal{A}}\frac{\partial v(w_A,a_A)}{\partial w_A} + \frac{-2\lambda a[1+\lambda(1-a)]+\lambda(1+\lambda-\lambda a^2)}{2}\frac{\partial v(w_A,a_A)}{\partial a_A}\right) \\ 0 &= \lambda w_{\mathcal{A}}\frac{\partial v(w_A,a_A)}{\partial w_A} - \lambda\frac{(1+\lambda)(2a-1)-\lambda a^2}{2}\frac{\partial v(w_A,a_A)}{\partial a_A}. \end{aligned} \tag{8}$$

To show that a^* is a unique solution, I show that $\frac{d^2 v(w_A,a_A)}{da^2} < 0$ for any $a > a^*$. It is easily verified that total derivative of the right hand side of equation 8 is negative, that is:

$$\begin{aligned} &\lambda w_{\mathcal{A}}\left(\frac{\partial^2 v(w_A,a_A)}{\partial^2 w_A}\frac{\partial w_A(a)}{\partial a} + \frac{\partial^2 v(w_A,a_A)}{\partial w_A \partial a_A}\frac{\partial a_A(a)}{\partial a}\right) - \lambda[(1+\lambda)-\lambda a]\frac{\partial v(w_A,a_A)}{\partial a_A} \\ &-\lambda\frac{(1+\lambda)(2a-1)-\lambda a^2}{2}\left(\frac{\partial^2 v(w_A,a_A)}{\partial w_A \partial a_A}\frac{\partial w_A(a)}{\partial a} + \frac{\partial^2 v(w_A,a_A)}{\partial^2 a_A}\frac{\partial a_A(a)}{\partial a}\right) < 0. \end{aligned}$$

The first term inside the first parenthesis is negative because $v_{ww} < 0$ and $w'_A(a) > 0 \ \forall a \in [0,1]$ by assumption. The second term inside the parenthesis is negative because $v_{wa} \geq 0$ by assumption, and $a'_A(a)$ must be negative in order for equation 7 to hold. The third term is negative because the partial derivatives of $v(w,a)$ are positive. Expression $-\lambda\frac{(1+\lambda)(2a-1)-\lambda a^2}{2}$ is negative if equation 8 holds. So it suffices to show that the two terms inside the last parenthesis are positive. The first term is positive because v_{wa} is positive by assumption and $w'_A(a) > 0 \ \forall a \in [0,1]$, and the second is positive because $v_{aa} < 0$ by assumption and $a'_A(a)$ must be negative in order for equation 7 to hold. Hence, a^* is unique.

It follows that the optimal discrimination value for $\mathcal{A}$ is $d^* = d(a^*)$. As argued above, there exists $\lambda' > 0$ such that for any $\lambda \in [0,\lambda')$, $d(a)$ is a strictly increasing function; hence, if $\lambda < \lambda'$, then d^* is unique. Note that $\frac{dv(\omega_A,a_A)}{da} > 0$ for any $a \leq \frac{1}{2}$, hence in order to satisfy the first order condition, it must be that $a^* > \frac{1}{2}$, and since it has already been established that the solution is interior, it follows $a^* \in (\frac{1}{2},1)$ and $d^* = d(a^*) > 0$ as claimed. ■

Proposition 2

Proof. Note that

$$w_A(a) = \frac{w_{\mathcal{A}} + \lambda(1-a)w_{\mathcal{D}}}{1+\lambda(1-a)} \text{ and} \tag{9}$$

$$\frac{\partial w_A}{\partial a} = \frac{-\lambda w_{\mathcal{D}}[1+\lambda(1-a)] + [w_{\mathcal{A}} + \lambda(1-a)w_{\mathcal{D}}]\lambda}{[1+\lambda(1-a)]^2} = \frac{\lambda(w_{\mathcal{A}} - w_{\mathcal{D}})}{[1+\lambda(1-a)]^2}, \tag{10}$$

so the first order condition is

$$\frac{dv(w_A, a_A)}{da} = \frac{\partial w_A}{\partial a}\frac{\partial v(w_A, a_A)}{\partial w_A} + \frac{\partial a_A}{\partial a}\frac{\partial v(w_A, a_A)}{\partial a_A} = 0,$$

which implies (compare to equation 8 in the proof of proposition 1):

$$0 = \lambda(w_{\mathcal{A}} - w_{\mathcal{D}})\frac{\partial v(w_A, a_A)}{\partial w_A} - \lambda\frac{(1+\lambda)(2a-1) - \lambda a^2}{2}\frac{\partial v(w_A, a_A)}{\partial a_A}. \tag{11}$$

Given a fixed $w_{\mathcal{A}}$, if $w_{\mathcal{D}}$ increases, the first term in equation 11 decreases; the second term must then increase for the equality to hold. For a sufficiently small λ, the second term is decreasing in a, so $a^*(w_{\mathcal{A}}, w_{\mathcal{D}})$ is decreasing in $w_{\mathcal{D}}$. As shown in the proof of proposition 1 for the case $w_{\mathcal{D}} = 0$, if λ and the wealth gap are sufficiently small, $d(a)$ is strictly increasing in a. Generalize the notation to let $d(a, w_{\mathcal{A}}, w_{\mathcal{D}})$ denote the level of difficulty that makes $i \in \mathcal{D}_E$ with ability $a_i = a$ indifferent between assimilation or not, as a function of both wealth levels. Since $w_{\mathcal{D}} = 0$ was merely a normalization, if λ and $w_{\mathcal{A}} - w_{\mathcal{D}}$ are sufficiently small, by the same argument $d(a, w_{\mathcal{A}}, w_{\mathcal{D}})$ is increasing in a. For any $w_1 > w_0$, $d(a, w_{\mathcal{A}}, w_{\mathcal{D}})|_{w_{\mathcal{D}}=w_1} < d(a, w_{\mathcal{A}}, w_{\mathcal{D}})|_{w_{\mathcal{D}}=w_0}$ because, given a fixed $w_{\mathcal{A}}$, the incentive to assimilate is lower if $w_{\mathcal{D}}$ is higher. Thus,

$$d(a, w_{\mathcal{A}}, w_{\mathcal{D}})|_{a=a^*(w_{\mathcal{A}},w_1), w_{\mathcal{D}}=w_1} < d(a, w_{\mathcal{A}}, w_{\mathcal{D}})|_{a=a^*(w_{\mathcal{A}},w_0), w_{\mathcal{D}}=w_1} < d(a, w_{\mathcal{A}}, w_{\mathcal{D}})|_{a=a^*(w_{\mathcal{A}},w_0), w_{\mathcal{D}}=w_0}$$

so $d^*(w_{\mathcal{A}}, w_{\mathcal{D}})$ is strictly decreasing in $w_{\mathcal{D}}$.

Similarly, for the second part of the proposition, given any sufficiently small fixed wealth gap Δ, if $w_{\mathcal{A}}$ and $w_{\mathcal{D}}$ increase in the same quantity, then $\frac{\partial v(w_A, a_A)}{\partial w_A}$ decreases by assumption, so the first term of the summation in equation 11 decreases. The rest of the argument is analogous to the case in the first part of the proposition.

The third statement notes that if the marginal utility of wealth converges to zero, then for a sufficiently high level of wealth, which depends on the wealth gap, the argument in the second statement applies. ■

Claim 3

Proof. Suppose $d_R > d_R^-$. Given d_R, choose $\tilde{d}$ to maximize the utility of the advantaged agents. Let $a(d, d_R, \xi_i)$ be the ability level of $i \in \mathcal{D}_E$ who is indifferent between assimilating or not given d and d_R and given that i draws an idiosyncratic cost ξ_i.

Advantaged agents want agents in $\mathcal{D}_E$ to assimilate if and only if $a_i \geq a^*$, but no $i \in \mathcal{D}_E$ with ability $a_i < a(d, d_R, \xi)$ assimilates. Consider the pair $(\hat{d}, d_R^-)$ such that the fraction of agents in $\mathcal{D}_E$ who assimilate given these difficulty levels is the same as with $(\tilde{d}, d_R)$. This implies that $\hat{d} > \tilde{d}$. Let $a^=$ be such that

$$d_R - d_R^- = (\hat{d} - \tilde{d})c(a^=).$$

Agent i with $a_i = a^=$ is equally likely to assimilate given $(\tilde{d}, d_R)$ or $(\hat{d}, d_R^-)$. Agents with $a_i > a^=$ are more likely and agents with $a_i < a^=$ less likely to assimilate given $(\hat{d}, d_R^-)$. Thus, given $(\hat{d}, d_R^-)$, a_A is higher than under $(\tilde{d}, d_R)$, and w_A is the same by assumption, hence the representative agent who chooses d and d_R is better off choosing $(\hat{d}, d_R^-)$. It follows that it is out of equilibrium to choose $d_R > d_R^-$. ■

Proposition 4

Proof. By an analogous argument as in the proof of proposition 1, there is a cutoff $a(d,\Theta) \in [0,1]$ such that for any d and Θ, at stage 3 an agent $k \in \mathcal{D}_E$ chooses $e_k = 1$ if $a_k > a(d,\Theta)$ and chooses $e_k = 0$ if $a_k \leq a(d,\Theta)$, and again by an analogous argument, if λ is low enough, then the solution to the joint assimilation decision is unique, which implies that if the distribution of ability in $\mathcal{D}_E$ has positive density everywhere in $[0,1]$, then $a(d,\Theta)$ is uniquely defined. If the distribution of ability in $\mathcal{D}_E$ assigns zero density to ability values in some non empty open interval (a_1, a_2) and the solution to the assimilation problem is such that $k \in \mathcal{D}_E$ chooses $e_k = 1$ if $a_k \geq a_2$ and chooses $e_k = 0$ if $a_k \leq a_1$, then any value $a(d,\Theta) \in [a_1, a_2)$ is a valid cutoff. In this case let $a_1(d,\Theta) = a_1$ denote the lowest possible cutoff and the highest ability among those who choose not to assimilate, and let $a_2(d,\Theta) = a_2$ denote the lowest ability among those who assimilate.

Anticipating the equilibrium of the subgame at stage 3, the unique best response at stage 2 of any agent k not subject to peer pressure is $x_k = \theta_k$. Choosing a lower level of education decreases $\psi(w_k, a_k)$, without increasing $u_k(d,e)$. Given $\mathcal{J}_E \in \{\mathcal{A}_E, \mathcal{D}_E\}$, an agent $k \in \mathcal{J}_E$ chooses $x_k \in \{\theta_k, x^P_{\mathcal{J}}\}$. Suppose $\theta_k \leq x^P_{\mathcal{J}}$. Then choosing an education below θ_k decreases $\psi(w_k, a_k)$ without having any other effect. Suppose $x^P_{\mathcal{J}} < \theta_k$. Then choosing $x_k = \theta_k$ dominates choosing any $x_k \in (x^P_{\mathcal{J}}, \theta_k)$ and choosing $x_k = x^P_{\mathcal{J}}$ dominates choosing any $x_k < x^P_{\mathcal{J}}$. Hence $x_k \in \{\theta_k, x^P_{\mathcal{J}}\}$. Between these two values there is a trade-off. Choosing θ_k brings a direct advantage of approximately $\psi(w_k, \theta_k) - \psi(w_k, x^P_{\mathcal{J}})$, where the exact value of the difference depends on realization of the noise term ε, and an indirect advantage of assimilating at a lower cost if k assimilates; whereas, choosing θ_k brings a cost K. The benefit of choosing θ_k over $x^P_{\mathcal{J}}$ is strictly increasing in θ_k, starting at a value of zero for $\theta_k = x^P_{\mathcal{J}}$. The cost is constant. Hence, there is a cutoff $\theta(K, x^P_{\mathcal{J}})$ increasing in both terms such that agents above the cutoff choose $x_k = \theta_k$ and incur the penalty K and agents below the cutoff avoid the penalty by choosing $x_k = x^P_{\mathcal{J}}$.

Recall $j \in \mathcal{D}$ with $\theta_j \leq \frac{1}{2}$ is the representative agent who chooses $x^P_{\mathcal{D}}$, and $i \in \mathcal{A}$ is the representative agent who chooses $x^P_{\mathcal{A}}$ and d.

Given the equilibrium strategies at stages 2 and 3, suppose that $(x^P_{\mathcal{D}}, (x^P_{\mathcal{A}}, d))$ are mutual best responses at stage 1. Let $\Theta(x^P_{\mathcal{D}}, (x^P_{\mathcal{A}}, d))$ be the limit as $\varepsilon \to 0$ of distribution of ability levels that occurs in equilibrium given $(x^P_{\mathcal{D}}, (x^P_{\mathcal{A}}, d))$ and the equilibrium actions at stages 2 and 3. Let $\alpha \equiv a(d, \Theta(x^P_{\mathcal{D}}, (x^P_{\mathcal{A}}, d)))$, so that $k \in \mathcal{D}_E$ assimilates if and only if $a_k > \alpha$. Suppose $x^P_{\mathcal{A}} < 1$. Then there exists $\tilde{d} > d$ such that $a(\tilde{d}, \Theta(x^P_{\mathcal{D}}, (1, \tilde{d}))) = \alpha$. Increasing the peer punishing threshold from $x^P_{\mathcal{A}}$ to 1 increases $a_{\mathcal{A}}$, making assimilation more attractive, but increasing the difficulty of assimilation reduces the incentives to assimilate, so there exist some value $\tilde{d} > d$ that exactly compensates the increase in $a_{\mathcal{A}}$ so that the same agent is indifferent about assimilation given $(x^P_{\mathcal{D}}, (x^P_{\mathcal{A}}, d))$ or given $(x^P_{\mathcal{D}}, (1, \tilde{d}))$. But i strictly prefers the second case: the set of agents who assimilate (and their ability level) is the same and the ability of advantaged agents $a_{\mathcal{A}}$ is higher under $(x^P_{\mathcal{D}}, (1, \tilde{d}))$; it follows that a_A is higher. Hence $x^P_{\mathcal{A}} < 1$ is not a best response, and hence in equilibrium $x^P_{\mathcal{A}} = 1$. Let $a_J(a)$ be the average ability in social group $J \in \{A, D\}$ given that every $k \in \mathcal{D}_E$ assimilates if and only if $a_k > a$. Let $w_A(a)$ be the average wealth in group A given that every $k \in \mathcal{D}_E$ assimilates if and only if $a_k > a$. Let $\alpha_A(a)$ be such that $v(w_{\mathcal{D}}, \alpha_A(a)) = v(w_A(a), \alpha_A(a))$. Then social group A is indifferent about the assimilation decision of an agent with ability $\alpha_A(a)$.

I show that $(x^P_{\mathcal{D}}, (1, d))$ are mutual best responses at stage 1 if and only if the following conditions are satisfied:

i) $a_D(a(d, \Theta(x^P_{\mathcal{D}}, (1, d)))) \leq x^P_{\mathcal{D}}$

ii) $x^P_{\mathcal{D}} \leq \alpha_A(a(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$

iii) $\alpha_A(a(d, \Theta(x^P_{\mathcal{D}}, (1, d)))) \leq a_2(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$.

iv) An agent with ability $x^P_{\mathcal{D}}$ is indifferent about assimilation.

Condition i) means that $j \in \mathcal{D}$ prefers agents with ability $x^P_{\mathcal{D}}$ to not assimilate. Condition ii) means that $i \in \mathcal{A}$ prefers them to not assimilate. Condition iii) says that i strictly prefers those who assimilate to indeed assimilate.

If i) is violated, j is better off deviating to $\tilde{x}^P_{\mathcal{D}} > x^P_{\mathcal{D}}$ which increases the ability in group D by letting the peer pressured agents acquire greater ability, which in this case benefits D even if it leads the agents to assimilate.

If ii) is violated, i can profitably deviate by lowering d to make the agents with ability $x^P_{\mathcal{D}}$ assimilate.

If iii) is violated, i can profitably deviate by increasing d enough to prevent the assimilation of agents with ability $a_2(d, \Theta(x^P_{\mathcal{D}}, (1, d)))$.

If iv) is violated and an agent with ability $x^P_{\mathcal{D}}$ strictly prefers not to assimilate, then j can deviate to $\tilde{x}^P_{\mathcal{D}} > x^P_{\mathcal{D}}$ and increase the equilibrium average ability in group D. If an agent with type $x^P_{\mathcal{D}}$ strictly prefers to assimilate, then j can deviate to $\tilde{x}^P_{\mathcal{D}} < x^P_{\mathcal{D}}$ low enough to prevent the assimilation of agents with ability $a(d, \Theta(\tilde{x}^P_{\mathcal{D}}, (1, d)))$. Since $a(d, \Theta(\tilde{x}^P_{\mathcal{D}}, (1, d))) - a(d, \Theta(x^P_{\mathcal{D}}, (1, d)))$ converges to zero as $\lambda \to 0$; since in order for d to be a best response it must in this case be that $\alpha_A(a(d, \Theta(x^P_{\mathcal{D}}, (1, d)))) \leq a(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$; and since $\alpha_A(a(d, \Theta(x^P_{\mathcal{D}}, (1, d)))) > a_D(a(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$, it follows that for a small enough λ, $a(d, \Theta(\tilde{x}^P_{\mathcal{D}}, (1, d))) > a_D(a(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$ and j deviates profitably by setting the threshold for peer punishments at the lower level $\tilde{x}^P_{\mathcal{D}}$ that prevents some agents from assimilating.

If all four conditions are satisfied, j cannot profitably deviate by setting a punishing threshold below $x^P_{\mathcal{D}}$, because this causes agents who stay in D to acquire a lower ability, and j cannot profitably deviate by setting a higher threshold because this causes the agents with ability $x^P_{\mathcal{D}}$ to assimilate (by condition iv), which hurts j (by condition i). Agent i cannot profitably deviate by setting a lower d because this causes the agents with ability $x^P_{\mathcal{D}}$ to assimilate, which is detrimental to j (by condition ii) and j cannot profitably deviate to a higher d because this either has no effect, or it causes the agents with ability $a_2(d, \Theta(x^P_{\mathcal{D}}, (1, d)))$ not to assimilate (by condition iii).

It remains to be shown that all conditions are satisfied for some value of $x^P_{\mathcal{D}}$ and d. Condition iv pins a unique value of d strictly increasing in $x^P_{\mathcal{D}}$ for any given $x^P_{\mathcal{D}}$. Hence, it suffices to show that conditions i-iii are simultaneously satisfied for some $x^P_{\mathcal{D}} < 1$. Condition i sets a lower bound on $x^P_{\mathcal{D}}$. So does condition iii. Condition ii sets up an upper bound below one. Since $a_D(a(d, \Theta(x^P_{\mathcal{D}}, (1, d)))) < \alpha_A(a(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$ for any $x^P_{\mathcal{D}}$, the upper bound implicit in condition ii is strictly higher than the lower bound implicit in condition i. For a small enough wealth differential $w_{\mathcal{A}} - w_{\mathcal{D}}$, $\alpha_A(a(d, \Theta(x^P_{\mathcal{D}}, (1, d))))$ is bounded below 1; whereas, $a_2(d, \Theta(x^P_{\mathcal{D}}, (1, d))) > x^P_{\mathcal{D}}$ for any $x^P_{\mathcal{D}}$ and $a_2(d, \Theta(x^P_{\mathcal{D}}, (1, d)))$ is continuously and strictly increasing in $x^P_{\mathcal{D}}$, hence there exist $x_1, x_2 \in (0, 1)$ such that $x_1 < x_2$ and $\alpha_A(a(d, \Theta(x_1, (1, d)))) = a_2(d, \Theta(x^P_{\mathcal{D}}, (1, d)))$ and $\alpha_A(a(d, \Theta(x_2, (1, d)))) = x_2$. For $x^P_{\mathcal{D}} \in [x_1, x_2]$, conditions ii and iii are satisfied. Thus, there exist values for which all four conditions are satisfied, equilibria exists and in all equilibria, $x^P_{\mathcal{D}}$ is bounded away from one. ■

References

[1] George A. Akerlof and Rachel E. Kranton. Economics and Identity. *Quarterly Journal of Economics*, 115(3):715–753, 2000.

[2] David Austen-Smith and Roland G. Fryer. An Economic Analysis of Acting White. *Quarterly Journal of Economics*, 120(2):551–583, 2005.

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

[4] Alberto Bisin and Thierry Verdier. The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97(2):298–319, 2001.

[5] Roland Bénabou and Jean Tirole. Identity, Morals and Taboos: Beliefs as Assets. *Quarterly Journal of Economics*, pages –, forthcoming.

[6] Randall L. Calvert. *Political Science: The State of the Discipline, 3rd ed.*, chapter Rationality, Identity and Expression. Eds. I. Katznelson and H. Milner, W.W. Norton and APSA, 2002.

[7] Pew Research Center. A year after obama's election. blacks upbeat about black progress, prospects. Survey, Jan 2010.

[8] James S. Coleman. *The Adolescent Society.* Free Press New York, 1961.

[9] Sergio Currarini, Matthew O. Jackson, and Paolo Pin. An Economic Model of Friendship: Homophily, Minorities and Segregation. *Econometrica*, 77(4):1003–1045, 2009.

[10] Sergio Currarini, Matthew O. Jackson, and Paolo Pin. Identifying Sources of Race-Based Choice and Change in High School Friendship Network Formation. *Proceedings of the National Academy of Science*, 107:4857–4861, 2010.

[11] Joan de Marti and Yves Zenou. Ethnic identity and social distance. CEPR discussion paper 7566, 2009.

[12] Federico Echenique, Roland G. Fryer, and Alex Kaufman. Is School Segregation Good or Bad? *American Economic Review*, 96(2):265–269, may 2006.

[13] James D. Fearon and David D. Laitin. Violence and the Social Construction of Ethnic Identity. *International Organization*, 54:845–877, 2000.

[14] Raymond Fisman, Sheena S. Iyengar, Emir Kamenica, and Itamar Simonson. Racial Preferences in Dating. *Review of Economic Studies*, 75(1):117–132, 2008.

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

[16] Signithia Fordham. *Blacked Out: Dilemmas of Race, Identity and the Success at Capital High*. Chicago, University of Chicago Press, 1996.

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

[18] Roland Fryer. *Handbook of Social Economics*, chapter The Importance of Segregation, Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap. Eds. Jess Benhabib, Alberto Bisin and Matthew Jackson. Elsevier, 2010.

[19] Roland D. Fryer. Acting White. *Education Next*, 6(1):–, 2006.

[20] Roland D. Fryer. A Model of Social Interactions and Endogenous Poverty Traps. *Rationality and Society*, 19(3):335–366, 2007.

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

[22] Roland D. Fryer and Paul Torelli. An empirical analysis of 'acting white'. *Journal of Public Economics*, 94(5-6):380–396, 2010.

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

[24] Claire A. Hill. The Law and Economics of Identity. *Queen's Law Journal*, 32(2):389–445, 2007.

[25] Karla Hoff and Arijit Sen. *Poverty Traps*, chapter The Kin System as a Poverty Trap? S. Bowles, S. Durlauf, K. Hoff eds. Princeton U. Press, 2006.

[26] Michael A. Hogg. *Handbook of Self and Identity*, chapter Social Identity. Guilford Press, 2003.

[27] Michael A. Hogg and Deborah J. Terry. Social Identity and Self-Categorization Processes in Organizational Contexts. *Academy of Management Review*, 25(1):121–140, 2000.

[28] Lu Hong and Scott E. Page. Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004.

[29] Daniel P. Hopkins. Politicized Places: Explaining Where and When Immigrants Provoke Local Opposition. *American Political Science Review*, 104(1):40–60, 2010.

[30] Richard Jenkins. *Social Identity*. Routledge, London, 1996.

[31] Desmond King. *Making Americans*. Harvard University Press, Cambridge, 2002.

[32] Esteban F. Klor and Moses Shayo. Social Identity and Preferences over Redistribution. *Journal of Public Economics*, 94:269–278, 2010.

[33] Edward P. Lazear. Culture and Language. *Journal of Political Economy*, 107(6):S95–S126, 1995.

[34] John H. McWhorter. *Losing the Race: Self-Sabotage in Black America*. New York, Free Press, 2000.

[35] Eleonora Patacchini and Yves Zenou. Racial identity and education. IZA discussion paper 2046, 2006.

[36] Elizabeth M. Penn. Citizenship versus Ethnicity: The Role of Institutions in Shaping Identity Choice. *Journal of Politics*, 70(4):956–973, 2008.

[37] Robert D. Putnam. E Pluribus Unum: Diversity and Community in the Twenty-First Century. *Scandinavian Political Studies*, 30(2):137–174, 2007.

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

[39] Robert E. Slavin. Can Financial Incentives Enhance Educational Outcomes? *Educational Research Review*, 5:68–80, 2010.

[40] Paul M. Sniderman, Louk Hagendoorn, and Markus Prior. Predisposing Factors and Situational Triggers. *American Political Science Review*, 98(1):35–49, 2004.

[41] Michael Spence. Job Market Signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.