Discrimination and Assimilation*

Jon X Eguia†
New York University
April 25, 2012

Abstract

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

*I am grateful to Will Terry for extensive comments, to Drew Conway for excellent research assistance, to Renee Bowen, Autumn Carter, Gary Cox, Eric Dickson, Oeindrila Dube, Raquel Fernandez, Rachel Kranton, Dimitri Landa, Maggie Penn, Ryan Pevnick, Carlo Prato, Ron Rogowski, Jakub Steiner, Leonard Wantchekon and attendants at talks at LSE, Northwestern, Stanford-GSB, Chicago, Harvard/MIT and NYU for their suggestions, and to the Ford Center for Global Citizenship and the Center for Mathematical Studies at the Kellogg School of Management for financial support during the academic year 2010-11.

†Email: eguia@nyu.edu. Mail: 19 West 4th St, 2nd floor, Dept. Politics, NYU. New York, NY 10012.
“When in Rome, do as the Romans do” (St. Ambrose, bishop of Milan, 384 AD).

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

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

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

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

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

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

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

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

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

¹If first generation immigrants do not assimilate, later generations of individuals brought up in the culture of an ethnic minority and not in the prevalent culture of their land of residence, such as Turks in Germany, or Hispanics and other minorities in the US (King 2002) face a qualitatively similar choice.
The cost of assimilation depends crucially on the attitude of the social group that the immigrant seeks to join. Hopkins (2010) identifies conditions that make a community more likely to be hostile to immigration. Sniderman, Hagendoorn and Prior (2004) find that Dutch citizens favor immigration by highly educated workers, and not by those who are only suited for unskilled jobs. Hainmueller and Hiscox (2010) refine this finding, distinguishing not only which immigrants inspire more negative reactions, but also which citizens (rich or poor) are more favorable toward each set of immigrants. They find that rich and poor US citizens alike strongly prefer high-skilled immigration and are opposed to low-skilled immigration. They review different economic theories of attitudes toward immigration\(^2\) and conclude that none explains their findings: “economic self-interest, at least as currently theorized, does not explain voter attitudes toward immigration.”

The theory I present in this paper is fully consistent with Hainmueller and Hiscox’s (2010) results: Economic self-interest leads low-skilled and high-skilled citizens alike to only welcome assimilation by high-skilled agents. While immigration leads to discrimination and social tensions as the native community seeks to deter many immigrants from assimilating, the successful integration of the most able immigrants ultimately results in a creative and intellectual boom for the community (Putnam 2007).

This paper builds upon an extensive literature on theories of social identity formation, and empirical and theoretical work on interethnic relationships. For interdisciplinary perspectives on identity, see Hogg and Terry (2000) and Hogg (2003) for social psychology; and the surveys in law and economics by Hill (2007), and in all of the social sciences by Jenkins (1996). Of particular relevance is the literature on identity economics, which argues that minorities adopt and pass on to their descendents identities that are anti-achievement (Akerlof and Kranton 2000), traditional (Bénabou and Tirole 2011) or ethnic (Bisin and Verdier 2000 and 2001) because if they shed this identity and embrace the productive/modern/majority identity, they suffer an exogenously given cost.

One question that I address in this paper is whether or not it is optimal for agents with a disadvantaged background to assimilate. Identity economics theories teach us that given a sufficiently high exogenous cost of assimilation, it is not. These theories do not ask a second question that motivates my research: why do agents with an advantaged background discriminate against those

\(^2\)These theories are based on labor market competition (Becker 1957, Scheve and Slaughter 2001) or on the burden on public services (Hanson, Scheve, and Slaughter 2007).
who seek to assimilate? I propose a theory which recognizes that the difficulty of assimilation is endogenous: it depends on the actions of the agents with an advantaged background, who choose their actions optimally to suit their own selfish interests.

Shayo (2009) presents a general framework in which the utility of an agent depends on the status of the group she identifies with, and on the distance in traits from the individual to the average member of the group. This distance depends on the actions of the agents. While Shayo (2009) does not solve the general model, his framework has proved useful in applications to redistributive policies (Klor and Shayo 2010) and institutional design in an ethnically divided society (Penn 2008). My theory departs from Shayo (2009) in various 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. I 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.

Research that focuses on behavior more than on internal notion of self seeks to identify conditions that leads agents to learn a common language (Lazear 1999), to form friendships (Currarini, Jackson and Pin 2009 and 2010; Fong and Isajiw 2000; Echenique, Fryer and Kaufman 2006; Pat-acchini and Zenou 2006; Marti and Zenou 2009), to go on dates (Fisman, Iyengar, Kamenica, and Simonson 2008) and to marry (Fryer 2007b) across ethnicities and races. The focus is on behavior and interactions with others, not on an introspective concept of self.

The closest reference to the my paper is Fryer’s (2007a) theory of endogenous group choice. Agents face a repeated choice between a narrow (disadvantaged) group, and a wider society. Each agent chooses to invest in skills that are group specific, or in skills that are valued by society at large. Members of the group reward the accumulation of group specific skills by greater cooperation with the agent. The theory in this paper and Fryer’s (2007a) reach a common conclusion: Disadvantaged agents suffer pressure from their peers to acquire a lower level of human capital.

The crucial difference between the theories is that Fryer’s (2007a) models an infinitely repeated game, which has multiple equilibria under standard folk theorem arguments. Different equilibria

---

3 Friendships, dates and marriages are all positive interactions. I study societies where the alternatives are assimilation and peaceful segregation. Societies where a more plausible alternative to assimilation is inter-ethnic conflict face a different strategic environment, discussed by Calvert (2002) and Fearon and Laitin (2000).
yield divergent (and outright contradictory) empirical implications; Fryer (2007a) describes the one equilibrium that supports his intuition. Whereas, I show that the relevant insight holds in every equilibrium in a simpler model. This model generates unambiguous empirical implications that are consistent with the previously unexplained findings by Hainmueller and Hiscox (2010) on attitudes toward immigration, and Fryer and Torelli (2010) on the acting white phenomenon.

The rest of the paper is organized as follows. First, I present the theory of assimilation, both without peer pressure (Proposition 1), and introducing peer pressure to explain the acting white phenomenon (Proposition 3). Second, I discuss how this theory fits available evidence better than alternative explanations, and I propose a policy intervention to align the incentives of minority students with the academic success of their co-ethnics.

1 Theory

Consider a society with a continuum of agents. Agents are distinguished by their background and their ability, both of which are exogenously given. The background of a half of the agents is advantaged. Let $A$ denote the set of agents with an advantaged background. Each agent $i \in A$ is endowed with wealth in quantity $w_A > 0$. The other half of the agents, denoted by $D$, have a disadvantaged background, and their endowment of wealth is $w_D = 0$. I interpret wealth very broadly, to include not only monetary or financial wealth, but also less tangible endowments such as status or social capital accumulated by members of the group.

Let $a_i$ denote the ability of agent $i$. Individual ability is private information. Assume that the distribution of ability among the set $A$ of agents with an advantaged background and the set $D$ of agents with a disadvantaged background is the same, with ability uniformly distributed in $[0, 1]$ in each set.

Agents choose their skill and their social group.

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

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

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

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

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

Any agent \( i \in \mathcal{D}_Y \) can choose to belong to \( D \) at no cost, or she can learn how to follow the norms
of the group \( A \) to then join \( A \), but this learning is costly. Let \( e_i \in \{0, 1\} \) be the choice of agent
\( i \in \mathcal{D}_Y \), 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 group \( A \), in which case I say that she assimilates.
Let \( e \) denote the decisions to assimilate by all agents in \( \mathcal{D}_Y \). Formally, \( e : [0, 1] \rightarrow \{0, 1\} \) is a
mapping from ability to assimilation decision. Given \( e \), the composition of the social groups is
\( A = A \cup \{i \in \mathcal{D}_Y : e_i = 1\} \) and \( D = D \setminus \{i \in \mathcal{D}_Y : e_i = 1\} \).

The cost of assimilating is \( e_i dc(s_i) \), where \( e_i \) acts as an indicator function making the cost zero if

\(^4 \)In the application of the theory to explain the acting white problem in subsection 1.2, I will interpret the set \( \mathcal{D}_Y \)
more precisely as the set of young agents with a disadvantaged background who attend desegregated schools.
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 $c(s_i)$ is a continuously differentiable, strictly positive, strictly decreasing function of skill $s_i$, which captures the intuition that more skilled agents can adapt at a lower cost.

The difficulty of assimilation $d$ is endogenous. It can be interpreted as the level of discrimination: If advantaged agents are welcoming to those who assimilate, $d$ is small. If the set of agents $A$ is hostile to those who do not master the cultural prerequisites of membership in $A$, then $d$ is high. Formally, I assume that an exogenously given finite subset $A_F \subset A$ of size $N$ of agents with an advantaged background chooses $d$.\footnote{We could let all agents with an advantaged background be involved in choosing $d$, but with an infinite number of agents, the strategic incentives to choose optimally vanish. Keeping the number finite generates strict incentives to choose optimally.} Size $N$ can be as small as one, or arbitrarily large. Label these agents according to their ability, so that $a_1 < a_2 < \ldots < a_N$. Each $i \in A_F$ strategically chooses $d_i \in \mathbb{R}_+$, and the vector $(d_1, \ldots, d_N)$ aggregates into a difficulty of assimilation $d \in \mathbb{R}_+$. I do not specify exactly how this aggregation takes place: it could be that the discrimination/difficulty faced by those who assimilate is the minimum of all the individual $d_i$ values, or the maximum, or the median, or any other order-statistic. I assume that for some integer $n \in \{0, 1, \ldots, N\}$, $d$ is the maximum real number such that at least $n$ agents in $A_F$ choose $d_i \geq d$. The intuition is that at least $n$ agents must wish to erect a given barrier to assimilation in order for this barrier to materialize.

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

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

where $U_i$ is the utility function of agent $i$, $\psi(w_i, s_i)$ is the direct utility that agent $i$ obtains from her exogenous endowment and her own skill, 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. The only assumptions on $\psi(w_i, s_i)$ are that it is continuous and strictly increasing in both arguments.

I assume that agents derive utility from the average wealth and skill 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 skilled 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 \( w_A \in \left[ \frac{w_A}{1+\lambda}, w_A \right] \), where the lower bound is achieved if every \( i \in D_Y \) assimilates, and the upper bound is achieved if none assimilate. The average wealth of agents in \( D \) is in any case \( w_A = 0 \). For any \( J \in \{ A, D, A, D \} \), let \( s_J \) be the average skill of agents in \( J \).

Let \( v(s_i, w_J, s_J) \) be the utility that an agent with skill \( s_i \) in social group \( J \in \{ A, D \} \) derives from the externalities coming from other agents in her group when the average wealth and skill of these agents are \( w_J \) and \( s_J \). Then, any \( i \in A \), who by assumption belongs to \( A \) at no cost, receives utility from externalities
\[
   u_i(d, e) = (1 - e_i)v(s_i, 0, s_A) + e_i[v(s_i, w_A, s_A) - dc(s_i)].
\]

(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, \( v(s_i, 0, s_D) \); whereas, if \( e_i = 1 \), agent \( i \) assimilates and attains \( u_i(d, e) = v(s_i, w_A, s_A) - dc(s_i) \). I assume that \( v \) is twice continuously differentiable, weakly increasing in \( s_i \) and strictly increasing in \( w_J \) and \( s_J \). For \( x, y \in \{ s_i, w_J, s_J \} \), let \( v_{xy} \) denote the cross-partial derivative with respect to \( x \) and \( y \). I assume that \( v_{w_J w_J} \leq 0 \) and \( v_{s_J s_J} \leq 0 \) (the marginal utility of externalities from average wealth and average skill is not increasing); \( v_{w_J s_J} \geq 0 \) (there is a complementarity between average group wealth and average group skill); and \( v_{s_i s_J} \geq v_{s_i w_J} = 0 \) (individual skill and group skill are complementary, while every member of a group equally enjoys the externality from the group’s average endowment).

### 1.1 Equilibrium Without Peer Pressure

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

First, each agents in \( A_F \subset A \) chooses her optimal discrimination level \( d_i \). These choices aggregate into a difficulty of assimilation \( d \), which becomes common knowledge.

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

Third, each agent \( i \in D_Y \) chooses whether or not to assimilate, \( e_i \in \{0, 1\} \). These choices determine the average skill and wealth of each social group, and hence payoffs.

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

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

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

\[
  s^*_i = \arg \max_{\{s\}} v(s_i, w_A(s), s_A(s)) \text{ s.t. } s_A(s) = \frac{1 + \lambda - \lambda s^2}{2 + 2\lambda(1 - s)} \text{ and } w_A(s) = \frac{w_A}{1 + \lambda(1 - s)},
\]

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

**Proposition 1** If the fraction \( \lambda \) of agents with disadvantaged background who have an opportunity to assimilate is sufficiently small, or if \( \log c(s_i) \) decreases sufficiently fast, then there exists a unique perfect Bayesian equilibrium, in which agents with a disadvantaged background choose to assimilate if and only if their ability is above \( a^* \).

Furthermore, if the difference in endowment \( w_A - w_D \) between agents with advantaged and disadvantaged background is not too large, then \( d^*_1 \geq d^*_2 \geq \ldots \geq d^*_{N-1} \geq d^*_N \) and \( a^* < 1 \) so that some agents assimilate.

A more technical statement of Proposition 1, along with its proof and all other proofs are in the Appendix.

The intuition of the result is that by setting an optimal (positive but not too large) difficulty of assimilation, agents with an advantaged background are able to optimally screen those who
assimilate and join their group: only agents with high ability (who in equilibrium are highly skilled) assimilate, and these are the agents who generate positive externalities. Agents with an advantaged background disagree on the optimal level of discrimination or difficulty of assimilation: because highly skilled individuals appreciate their group’s average skill more than less skilled individuals \((v_{s_i,s_j} \geq 0)\), highly skilled advantaged agents want to discriminate less (strictly less if \(v_{s_i,s_j} > 0\)) to assimilate more highly skilled agents with a disadvantaged background. Less skilled agents do not care as much for the increase in average skill that comes with assimilation of the more skilled disadvantaged agents, and they resent the decrease in average wealth endowment, so they prefer higher barriers to assimilation to let fewer agents assimilate. Only if the endowment gap is too large, all agents with an advantaged background agree that it is best not to let anyone assimilate. Otherwise the solution is interior, and the cutoff for assimilation maximizes the utility of one advantaged agent, the one who is pivotal in determining the level of discrimination.

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

An alternative sufficient condition is to assume that the cost of assimilation drops very rapidly (in relative terms) with skill. Formally, the condition is that \(\frac{c'(s)}{c(s)}\) be very negative. Intuitively, it means that the cost faced by a more skilled agent is only a small fraction of the cost paid by a less skilled agent. If agents with unequal ability face such different incentives, the equilibrium is unique separating agents with ability above or below the cutoff, regardless of the size \(\lambda\) of the set of agents who can assimilate.

On the other hand, if agents face more homogeneous costs and the set of agents who can assimilate is large, then for some functional forms a cascade may occur: once the most skilled agents with a disadvantaged background assimilate, the average skill among the agents remaining in the disadvantaged group may be so low that agents with intermediate skills face a greater incentive to assimilate as well. If so, advantaged agents are no longer able to optimally screen, and it can occur (examples are available from the author) that the advantaged agents set a very high \(d^*\) to forestall the cascading assimilation of too many agents, or there can be multiple equilibria.
depending on whether agents with a disadvantaged background coordinate to assimilate in very small or in very large numbers.

The model is a variation on the seminal signaling model by Spence (1973): Members of the set \(\mathcal{D}_Y\) choose whether to invest in assimilation techniques. This investment is less costly for more skilled agents, so in equilibrium the agents separate: Highly skilled agents invest, and less skilled agents do not. Observing the investments, the advantaged agents accept into their social group those who invested at least \(d^*\). Discrimination by means of imposing a cost of assimilation \(d^* > 0\) is a screening device that the advantaged agents use to separate agents with low skills from highly skilled agents.\(^6\)

I describe in the Appendix three generalizations to the model: (1) distinguishing between costs of assimilation based on behavioral norms that individuals must learn, and costs based on immutable exogenous traits such as race; (2) allowing for 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. I find that in order to provide optimal screening, discrimination must be based on malleable traits (culture, behavior, etc) and not on immutable traits that do not correlate with ability (skin color, place of birth, etc), and that qualitative results are robust if we allow for intrinsic preferences for or against diversity, or if we consider assimilation in both directions.

To study the comparative statics with changes in the endowment gap between groups, I relax the normalization that \(w_D = 0\), assuming instead that \(0 \leq w_D \leq w_A\), so that I can study the effect of increases in the wealth of each group independently. 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.

**Proposition 2** There exist \(\Delta > 0\) and \(\lambda' > 0\) such that for any wealth gap \(w_A - w_D \in (0, \Delta]\) and any \(\lambda < \lambda'\),

i) The equilibrium difficulty \(d^*\) and cutoff for assimilation \(a^*\) strictly decrease if \(w_D\) increases while \(w_A\) remains constant, and

\(^6\)By imposing a cost of assimilation, agents with an advantaged background both discriminate against all agents with a disadvantaged background, and—in a more favorable sense of the word— they discriminate among agents with a disadvantaged background, by discerning and distinguishing who are the most talented among them, who then assimilate.
ii) The equilibrium difficulty $d^*$ and cutoff for assimilation $a^*$ decrease if both $w_D$ and $w_A$ increase in the same amount.

The first result says that if the wealth gap is not too large, assimilation increases as agents with a disadvantaged background become wealthier and the endowment gap narrows. The second result notes that assimilation also increases if both groups become richer, keeping the endowment gap constant. The cutoffs strictly decrease if the externality from wealth has decreasing marginal utility.

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

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

### 1.2 Acting White Equilibrium with Peer Pressure

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

I interpret the choice of a skill level $s_i \in [0, a_i]$ as the choice to attain a level of success in school. Students who choose $s_i < a_i$ do not achieve their potential, come out of school with fewer skills, and are less able to succeed in society. Recall that the utility of agent $i$ is $\psi(w_i, s_i) + u_i(d, e)$. All else equal, every $i$ prefers the highest possible skill $s_i$ to maximize $\psi(w_i, s_i)$. But all else is not equal: In some schools, peers may punish those who excel.
I introduce peer pressure into the theory. Recall that the set $D_Y$ comprises the fraction $\lambda$ of agents with a disadvantaged background who choose strategically the social group they want to belong to. Think of them as young minority students who attend desegregated high schools. Assume that these agents are susceptible to peer pressure. For symmetry, assume as well that a set $A_Y \subset A$ of size $\lambda$ of young agents with an advantaged background are susceptible to peer pressure by other agents with an advantaged background.

I model peer pressure as follows: Let $m \in D$ and $l \in A$ be such that $a_i \leq \frac{1}{2}$ for $i \in \{m, l\}$. Agent $m$ chooses a skill threshold $s_P^D \in [0, 1]$ and agent $l$ chooses a skill threshold $s_P^A \in [0, 1]$. These levels become common knowledge. Every $i \in D_Y$ who chooses $s_i > s_P^D$ and every $i \in A_Y$ who chooses $s_i > s_P^A$ incur a fixed cost $K > 0$. I interpret this cost $K$ as a reduced form that captures the social cost of overachieving in school, which may manifest itself in punishments as physical bullying, or more mildly, in the form of social disaffection.\footnote{In practice, peer pressure must be implemented by a group. I do not seek to explain how groups solve coordination problems: what I seek to explain is how groups who can exert peer pressure use this tool optimally. Hence I blackbox the collective implementation of peer pressure, assuming that the cutoff is chosen by a single individual, and the cost $K$ incurred automatically. An extension showing that the equilibrium in Proposition 3 holds if punishments are determined by the aggregation of collective decisions is available from the author.}

Keeping all other assumptions on the direct utility function $\psi$, the cost function $c$ and the utility from externalities $v$ from the previous subsection, assume further that $v(s_i, w_J, s_J) = w_J + f(s_i)s_J$ for some arbitrary weakly increasing function $f$. The utility function of an agent $i \in D_Y$ is then:

\[
\psi(w_i, s_i) + (1 - e_i)f(s_i)s_D + e_i[w_A + f(s_i)s_A - dc(s_i)] - K \text{ if } s_i > s_P^D, \\
\psi(w_i, s_i) + (1 - e_i)f(s_i)s_D + e_i[w_A + f(s_i)s_A - dc(s_i)] \text{ if } s_i \leq s_P^D.
\]

The timing is as follows:

1. Agent $m \in D$ chooses a peer pressure threshold $s_P^D$. Simultaneously, agent $l \in A$ chooses the peer pressure threshold $s_P^A$ and the difficulty of assimilation $d$.\footnote{The main result is robust to variations in the timing of moves. For a model where $d$ is chosen before $s_P^D$ and $s_P^A$, see working paper versions of the manuscript from 2010. The equilibrium is also robust if different agents or sets of agents choose $s_P^A$ and $d$, but in this case I cannot rule out that other equilibria exist in which the agents with an advantaged background fail to coordinate resulting in jointly suboptimal $s_P^A$ and $d$.}

2. Each agent $i$ chooses her skill $s_i \in [0, a_i]$. For $J \in \{A, D\}$, any $k \in J_Y$ who chooses $s_i > s_P^J$ incurs punishment $K$.

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

At step 2, any agent $i \notin \mathcal{A}_Y \cup \mathcal{D}_Y$ chooses skill $s_i = a_i$. Any $i \in \mathcal{A}_Y$ chooses $s_i \in \{a_i, s_{A}^P\}$ and any $i \in \mathcal{D}_Y$ chooses $s_i \in \{a_i, s_{D}^P\}$.

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

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

**Proposition 3** For any $\gamma > \frac{1}{2}$ and any $\varepsilon > 0$, if $\lambda$ and $w$ are sufficiently small, an equilibrium in which $s_{D}^P < 1$ and $s_{A}^P = 1$ exists, and in any equilibrium $s_{D}^P < \gamma$ and $s_{A}^P > 1 - \varepsilon$.

Proposition 1 had shown that if the endowment of wealth and the set of young agents who can assimilate is not too large (or if the cost of assimilation decreases sufficiently rapidly with skill level), the equilibrium without peer pressure leads to assimilation, which harms agents with low ability and a disadvantaged background. Proposition 3 shows that these doubly disadvantaged agents respond optimally by punishing success in school. In all equilibria, highly able agents with a disadvantaged background are pressured to underperform; whereas, agents with an advantaged background are not subjected to the same pressure. This is the acting white phenomenon.

---

9 It is strictly dominated for these agents to choose $s_i < a_i$. We could assume directly that $s_i = a_i$ for any agent $i \notin \mathcal{A}_Y \cup \mathcal{D}_Y$ to let only young agents choose their skill. This assumption would be more consistent with the interpretation that $s_i$ measures accumulation of skills in school, and it would not change any result.
Note that in equilibrium, \( d \) is lower in the game with peer pressure than in the game without it: fewer highly skilled agents assimilate, and as a consequence, the average skill level in \( A \) is lower, so intergroup differences are smaller, making assimilation less desirable. I illustrate these and other differences with a numerical example.

**Example 4** Let \( w_A = 4 \), \( \lambda = 0.1 \), \( c(s) = \frac{1}{s_i} \), \( \psi(w_i, s_i) = w_i^{1/2} + 10s_i \), and \( v(s_i, w_J, s_J) = w_J^{1/2} + 10s_J \). Let \( U_A, U_{D-} \) and \( U_{D+} \) respectively denote the average utility of \( \{i \in A\} \), \( \{i \in D\} \) and \( \{i \in D: a_i \leq \frac{1}{2}\} \). Columns 2 and 3 compare the equilibrium outcomes under an assumption of no peer pressure (\( K = 0 \)) in column 2, and peer pressure (\( K = 1 \)) in column 3, where \( s_D^P = 0.6 \) and \( s_A^P = 1 \) are part of the equilibrium.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2) No peer pressure</th>
<th>(3) Peer pressure</th>
<th>(3)-(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d^* )</td>
<td>1.341</td>
<td>1.314</td>
<td>-0.027</td>
<td></td>
</tr>
<tr>
<td>( a^* )</td>
<td>0.610</td>
<td>0.676</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>( s_D )</td>
<td>0.487</td>
<td>0.488</td>
<td>+0.001</td>
<td></td>
</tr>
<tr>
<td>( U_A )</td>
<td>14.077</td>
<td>14.074</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td>( U_D )</td>
<td>9.896</td>
<td>9.800</td>
<td>-0.096</td>
<td></td>
</tr>
<tr>
<td>( U_{D-} )</td>
<td>7.376</td>
<td>7.384</td>
<td>+0.008</td>
<td></td>
</tr>
</tbody>
</table>

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

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

2 Discussion, Evidence and Policy Implications

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

The “oppositional culture” theory of Fordham and Ogbu (1986) and Fordham (1996) posits that academic failure is an integral component of African-American group identity: whites embrace values of studiousness and hard work, while minorities reject these values, embracing instead a counterculture defined in opposition to the mainstream values, in particular in opposition to the pursuit of success at school. They find that students in the 1980s perceived activities such as speaking standard English, getting good grades, or going to libraries as distinctly “white” and they stress that to engage in these behaviors is to give up membership in the black social group. They trace back the roots of black students’ self-identification with academic failure to a history of oppression in which whites (that is, society at large) negated their accomplishments regardless of their effort and objective merit.

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

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

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

The oppositional culture and the sabotage theories imply that the acting white problem ought to be more severe in schools with the least socioeconomic opportunities for upward mobility. The screening theory I have presented in this paper has the opposite empirical implication: the acting white phenomenon and the social price paid by the minority students who insist on achieving academic success should increase with the opportunities for upward mobility faced by the students.

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

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

The asymmetry across ethnic groups is the essence of the acting white phenomenon. In the screening theory I have developed, this asymmetry is obtained as a main result (Proposition 3), derived from primitives (agents’ utility functions, distribution of ability and technology for peer pressure) that are symmetric across groups, with the exception of wealth. Solely from an unequal endowment of wealth, it follows that agents with a disadvantaged background discourage their peers’s acquisition of skills, while agents with an advantaged background do not.\textsuperscript{10}

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

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

\textsuperscript{10}The symmetric assumption on the distribution of ability across groups is dynamically inconsistent with the equilibrium result, but relaxing this assumption would only reinforce the result and it would obscure the relevant insight that a disparity in the endowment of wealth suffices to drive the result.
Hispanic) students with lower grades build friendships mostly among their co-ethnics, which is fully consistent with the screening theory.

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

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

The punishment of high achieving African-American and Hispanic students is only an instance of a broader social phenomenon. In groups as diverse as the Buraku outcasts in Japan, Italian immigrants in Boston, the Maori in New Zealand and the working class in Britain, high-achievers have suffered a negative externality from their peer group (see Fryer 2007a for a discussion). Hoff and Sen (2006) report a strikingly similar problem in the context of informal insurance provided by extended families in the developing world: “If the kin group foresees that it will lose some of its most productive members as the economy opens up, it may take collective actions ex ante to erect exit barriers.” I interpret the acting white phenomenon as one such exit barrier.¹¹

¹¹ Religious doctrines opposing inter-faith marriage can also be understood as exit-deterrance strategies. A religion that sees assimilation as a threat can protect itself by condemning any interfaith marriage (as traditional Judaism does; see Deuteronomy 7:3-4); whereas, a religion that expects wives to be subordinate to their husbands maximizes its influence by prohibiting its women but allowing its men to marry outside the faith (as Islam does; see Quran 2:221 and Quran 5:5).
The screening theory’s external validity as an explanation not just of acting white, but of the broader phenomenon that underprivileged communities deter exit by making skill acquisition costly, is testable. Students in rural schools face an analogous strategic environment: academic success leads to migration to the city. Therefore, the theory predicts rural students who obtain top grades to be less popular, regardless of their race. In the United States, this can be tested using the Add Health dataset. An analogous prediction applies to other countries and contexts; in the words of Fryer and Torelli (2010): “any group presented with the same set of payoffs, strategies and so on, would behave identically.”

2.1 Policy Recommendation

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

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

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

I suggest instead that conditional 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 reward that is contingent on the total number of good 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 reward. I conjecture that under these incentives, the most able students who produce the public good enjoyed by all their classmates would no longer lose popularity for achieving the high grades that deliver these public goods.

3 Appendix

First I describe three generalizations to the model. A detailed formalization, and precise results with their proofs for these generalizations are available in the working paper of this manuscript, and from the author.

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

Different kinds of discrimination

In a society where the division of wealth corresponds to an ethnic divide of the population, agents in $D$ may differ from those in $A$ 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 agents with an advantaged background seek to harness the positive externalities provided by highly skilled individuals, an optimal discrimination policy must be based on an endogenous correlate of ability such as the ease of learning the arbitrary cultural norms of group $A$, rather than on an ascriptive characteristic that offers no information about the person’s skills.

Put it differently, even if advantaged agents care only about their self-interest and are unconcerned about the welfare of disadvantaged agents, as long as they are strategic, they do not discriminate on the basis of immutable characteristics such on skin color, race, place of birth.
Rather, strategic agents with an advantaged background prefer to screen on the basis of some observable characteristic that correlates with ability and skill. Agents with an advantaged background can construct and use a set of norms that are less costly to acquire for highly skilled agents, and then they can adopt a simple cut-off rule: Agents with a disadvantaged background who acquire a sufficiently high proficiency in the set of norms of $A$ must be very skilled, and thus they should be assimilated, while agents who do not acquire such ease with the chosen norms are rejected and not assimilated.

A qualification to this argument leads to the second generalization.

**Intrinsic preferences for or against diversity**

If agents have intrinsic preferences over exogenous attributes such as race or place of birth, they may prefer ceteris paribus to associate with those who look like them or come from the same town. The qualitative results in the theory are robust to these preferences: If agents in $A$ are prejudiced (Lupia 2011) or dislike some exogenous attribute of set $D$, agents in $A$ treat those in $D$ as if the wealth differential was higher, and as a result the equilibrium difficulty of assimilation $d^*$ rises and fewer agents assimilate. If agents in $D$ dislike some exogenous attribute of $A$, then agents in $D$ perceive the wealth difference as smaller, and the equilibrium difficulty of assimilation $d^*$ must be lower in order to entice agents with a disadvantaged background to assimilate. If both sets of agents dislike the exogenous attributes of the other set, then the effect on $d$ is ambiguous, but the number of agents who assimilate is smaller, resulting in voluntary segregation. Whereas, if ceteris paribus diversity increases agents’ payoffs (Hong and Page 2004), in equilibrium there is less discrimination and more assimilation.

**A symmetric society**

Consider a more symmetric strategic environment in which groups have different endowments that are not clearly ordered, and assimilation and discrimination occur in both directions. An interpretation of this symmetric version is that different agents have different priorities in life. Perhaps an economically disadvantaged group $D$ enjoys a greater artistic or musical richness in its community. Members of $D$ 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 $A$ 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 $i \in A$ is endowed with $w$ in quantity $w_A$.
and every $i \in \mathcal{D}$ is endowed with $m$ in quantity $m_{\mathcal{D}}$, while $w_{\mathcal{D}} = m_{\mathcal{A}} = 0$. 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 $\mathcal{A}$ at no cost, and if $i \in \mathcal{D}_Y$, then $i$ assimilates if and only if $s_i$ is sufficiently high. However, now assimilation goes both ways: Agent $i \in \mathcal{A}_Y$ who values $m$ assimilates into $\mathcal{D}$ if and only if she is sufficiently skilled.

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

Next I prove the results in this paper.

**Proposition 1** There exist $a^* > \frac{1}{2}$, $\bar{w} > 0$, $\bar{\lambda} > 0$ and $\bar{c} > 0$ such that if $\lambda < \bar{\lambda}$ or $-\bar{c} < c(s_i)$, then

a) there exists a unique perfect Bayesian equilibrium, in which any $i \in \mathcal{D}_Y$ with $a_i > a^*$ assimilates and any $i \in \mathcal{D}_Y$ with $a_i < a^*$ does not assimilate, and

b) if $w_{\mathcal{A}} < w'$, then in this equilibrium $0 \leq d_N^\ast \leq d_{N-1}^\ast \leq \ldots \leq d_2^\ast \leq d_1^\ast$ and $a^* < 1$.

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

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

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

$$w_A(s) = \frac{w_{\mathcal{A}}}{1 + \lambda(1 - s)},$$

$$s_A(s) = \left[\frac{1}{2} + \lambda(1 - s)\right]\frac{1}{2 + 2\lambda(1 - s)} = \frac{1 + \lambda - \lambda s^2}{2 + 2\lambda(1 - s)},$$

$$s_D(s) = \left[\frac{s}{2} + (1 - \lambda)(1 - s)\right]\frac{1}{s + (1 - \lambda)(1 - s)} = \frac{1 - \lambda + \lambda s^2}{2 - 2\lambda(1 - s)}.$$

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

Let d(s) be the value of d such that i ∈ D_Y with s_i = s is indifferent between assimilating or not given that other agents assimilating if and only if their skill is above s. This value is unique.

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

For any x, y, z ∈ R, let v(s_i, w_j, s_j)_{s_i=x, w_j=y, s_j=z} denote the value of v(s_i, w_j, s_j) evaluated at s_i = x, w_j = y and s_j = z. Then

\[
d(s) = \frac{v(s_i, w_j, s_j)_{s_i=s, w_j=w_A(s), s_j=s_A(s)} - v(s_i, w_j, s_j)_{s_i=s, w_j=0, s_j=s_D(s)}}{c(s)}.
\]

Note that if λ = 0, then

\[
d(s) = \frac{v(s_i, w_j, s_j)_{s_i=s, w_j=w_A(s), s_j=\frac{1}{2}} - v(s_i, w_j, s_j)_{s_i=s, w_j=0, s_j=\frac{1}{2}}}{c(s)},
\]

which is a strictly increasing, continuously differentiable function, with

\[
d'(s) = -\frac{c'(s)\left[v(s_i, w_j, s_j)_{s_i=s, w_j=w_A(s), s_j=\frac{1}{2}} - v(s_i, w_j, s_j)_{s_i=s, w_j=0, s_j=\frac{1}{2}}\right]}{[c(s)]^2} > 0.
\]

For any λ ∈ [0, 1), since w_A(s), s_A(s), s_D(s), c(s), c'(s) are continuous in λ for any λ ∈ [0, 1), v(s_i, w_j, s_j) is continuous, and c(s) is positive for any s, so both d(s) and d'(s) are continuous in λ for any λ ∈ [0, 1). Therefore, there exists λ' > 0 such that if λ < λ', then d'(s) > 0.

Alternatively, for any λ ∈ [0, 1],

\[
c(s)d'(s) = \left[\frac{d}{ds}v(s_i, w_j, s_j)_{s_i=s, w_j=w_A(s), s_j=s_A(s)} - \frac{d}{ds}v(s_i, w_j, s_j)_{s_i=s, w_j=0, s_j=s_D(s)}\right] - \frac{c'(s)}{c(s)}\left[v(s_i, w_j, s_j)_{s_i=s, w_j=w_A(s), s_j=s_A(s)} - v(s_i, w_j, s_j)_{s_i=s, w_j=0, s_j=s_D(s)}\right].
\]

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

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

Fourth Step: Find the optimal \(d^*_i\) for each \(i \in A_F\).

Let \(s^*(s_i) = \arg \max_{s \in [0,1]} v(s_i, w_J, s_J)\) s.t.
\[
\begin{align*}
    w_J &= w_A(s) = \frac{w_A}{1 + \lambda(1 - s)}, \\
    s_J &= s_A(s) = \frac{1 + \lambda - \lambda s^2}{2 + 2\lambda(1 - s)}.
\end{align*}
\]

Since \(v(s_i, w_A(s), s_A(s))\) is continuous in \(s\), it achieves a maximum on the compact set \([0,1]\), so a solution exists. I show that for a sufficiently low \(w_A\), the solution must be interior. First, \(s = 0\) is not a solution, because 
\[
\frac{dv(s_i, w_A, s_A)}{ds} > 0 \quad \text{at} \quad s = 0.
\]
Second, \(s = 1\) is not a solution for a low enough \(w_A\), because if \(s = 1\), then
\[
\frac{dv(s_i, w_A, s_A)}{ds} = \lambda w_A \frac{\partial v(s_i, w_A, s_A)}{\partial w_A} + \frac{-\lambda - \lambda^2}{2} \frac{\partial v(s_i, w_A, s_A)}{\partial s_A}
\]
which is negative if
\[
w_A < \frac{1 + \lambda}{2} \frac{\partial v(s_i, w_A, s_A)}{\partial s_A} \frac{\partial v(s_i, w_A, s_A)}{\partial w_A}.
\]

Since the solution is interior, it satisfies the first order condition
\[
\frac{dv(s_i, w_A, s_A)}{ds} = \frac{\partial w_A}{\partial s} \frac{\partial v(s_i, w_A, s_A)}{\partial w_A} + \frac{\partial s_A}{\partial s} \frac{\partial v(s_i, w_A, s_A)}{\partial s_A} = 0.
\]

Note that
\[
\frac{\partial w_A}{\partial s} = \frac{\lambda w_A}{[1 + \lambda (1 - s)]^2} \quad \text{and} \quad \frac{\partial s_A}{\partial s} = \frac{-2\lambda s[1 + \lambda (1 - s)] + \lambda(1 + \lambda - \lambda s^2)}{2[1 + \lambda (1 - s)]^2}.
\]
so a solution \( s = s^*(s_i) \) satisfies

\[
0 = \frac{1}{[1 + \lambda(1 - s)]^2} \left( \lambda w_A \frac{\partial v(s_i, w_A, s_A)}{\partial w_A} + \frac{-2\lambda s[1 + \lambda(1 - s)] + \lambda(1 + \lambda - \lambda s^2)}{2} \frac{\partial v(s_i, w_A, s_A)}{\partial s_A} \right)
\]

\[
0 = \lambda w_A \frac{\partial v(s_i, w_A, s_A)}{\partial w_A} - \lambda(1 + \lambda)(2s - 1) - \lambda s^2 \frac{\partial v(s_i, w_A, s_A)}{\partial s_A}. \tag{6}
\]

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

\[
\lambda w_A \left( \frac{\partial^2 v(s_i, w_A, s_A)}{\partial^2 w_A} w'_A(s) + \frac{\partial^2 v(s_i, w_A, s_A)}{\partial w_A \partial s_A} s'_A(s) \right) - \lambda(1 + \lambda) - \lambda s \frac{\partial v(s_i, w_A, s_A)}{\partial s_A} - \lambda(1 + \lambda)(2s - 1) - \lambda s^2 \left( \frac{\partial^2 v(s_i, w_A, s_A)}{\partial w_A \partial s_A} w'_A(s) + \frac{\partial^2 v(s_i, w_A, s_A)}{\partial^2 s_A} s'_A(s) \right) < 0.
\]

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

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

\[
\frac{\partial s_A \partial^2 v(s_i, w_A, s_A)}{\partial s_A \partial s_i} \leq 0
\]

hence an agent \( j \in A_F \) with \( s_J \geq s_i \) satisfies the first order equation 5 by setting \( s^*(s_j) \leq s^*(s_i) \) and thus \( d^*_j \leq d^*_i \).

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

**Proposition 2**

**Proof.** Note that

$$w_A(s) = \frac{w_A + \lambda(1 - s)w_D}{1 + \lambda(1 - s)} \text{ and }$$

$$\frac{\partial w_A}{\partial s} = \frac{-\lambda w_D[1 + \lambda(1 - s)] + [w_A + \lambda(1 - s)w_D]\lambda}{[1 + \lambda(1 - s)]^2} = \frac{\lambda(w_A - w_D)}{[1 + \lambda(1 - s)]^2},$$

so the first order condition is

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

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

$$0 = \lambda(w_A - w_D) \frac{\partial v(s_i, w_A, s_A)}{\partial w_A} - \lambda \frac{(1 + \lambda)(2s - 1) - \lambda s^2}{2} \frac{\partial v(s_i, w_A, s_A)}{\partial s_A}. \quad (9)$$

Given a fixed $w_A$, if $w_D$ increases, the first term in equation 9 decreases; the second term must then increase for the equality to hold. For a sufficiently small $\lambda$, the second term is decreasing in $s$, so for any $j \in A_F$, $s^*(s_j, w_A, w_D)$ is decreasing in $w_D$. As shown in the proof of proposition 1 for the case $w_D = 0$, if $\lambda$ and the wealth gap are sufficiently small, $d(s)$ is strictly increasing in $s$. Generalize the notation to let $d(s, w_A, w_D)$ denote the level of difficulty that makes $i \in D_Y$ with skill $s_i = s$ indifferent between assimilation or not, as a function of both wealth levels. Since $w_D = 0$ was merely a normalization, if $\lambda$ and $w_A - w_D$ are sufficiently small, by the same argument $d(s, w_A, w_D)$ is increasing in $s$. For any $w_1 > w_0$, $d(s, w_A, w_D)|_{w_D = w_1} < d(s, w_A, w_D)|_{w_D = w_0}$ because, given a fixed $w_A$, the incentive to assimilate is lower if $w_D$ is higher. Thus,

$$d(s, w_A, w_D)|_{s = s^*(s_j, w_A, w_1), w_D = w_1} < d(s, w_A, w_D)|_{s = s^*(s_j, w_A, w_0), w_D = w_1} < d(s, w_A, w_D)|_{s = s^*(s_j, w_A, w_0), w_D = w_0}$$
so \( d'(s_j, w_A, w_D) = d(s, w_A, w_D) |_{s=s^*(s_j, w_A, w_D)} \) is strictly decreasing in \( w_D \) for each \( j \in A_F \), and thus the equilibrium difficulty \( d'(w_A, w_D) \) is strictly decreasing in \( w_D \).

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

**Proposition 3** For any \( \gamma > \frac{1}{2} \) and any \( \varepsilon > 0 \), there exist \( \lambda(\gamma, \varepsilon) > 0 \) and \( w(\gamma, \varepsilon) > 0 \) such that for any \( \lambda < \lambda(\gamma, \varepsilon) \) and any \( w_A < w(\gamma, \varepsilon) \):

An equilibrium in which \( s_D^P < 1 = s_A^P \) exists, and in any equilibrium \( s_D^P < \gamma \) and \( s_A^P > 1 - \varepsilon \).

**Proof. Part 1:** I first prove the existence claim.

Let \( \Omega = (\Omega_{D_Y}, \Omega_{A_Y}) \) be the pair of distributive functions of the skill in \( D_Y \) and \( A_Y \).

Note first that at the second stage, for any \( i \in J_Y \) and \( J \in \{A, D\} \) choosing any \( s_i \notin \{a_i, s_J^P\} \) is strictly dominated either by \( s_i = a_i \) or by \( s_i = s_J^P \). Therefore, in equilibrium \( s_i \in \{s_i, s_J^P\} \) so that \( \Omega_{J_Y} \) has uniform density on \([0, s_J^P]\) and positive mass only at \( s_J^P \) (and it may or may not have positive density above \( s_J^P \) depending on the actions of the agents).

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

At the second stage, given \((d, s_D^P, s_A^P)\) and in anticipation of an equilibrium in stage 3, any \( i \notin D_Y \cup A_Y \) and any agent \( i \in J_Y \) with \( a_i < s_J^P \) uniquely best respond by choosing \( s_i = a_i \). Any agent \( i \in J_Y \) with \( a_i > s_J^P \) faces a trade-off: Choosing \( s_i = a_i > s_J^P \) she incurs a cost \( K \), but she derives a benefit in terms of direct utility \( \psi \) and in terms of a reduced cost of assimilation (if the assimilates). The benefit of choosing \( s_i = a_i \) is increasing in \( a_i \), while the cost is fixed at \( K \). Thus, there is a cutoff \( a(s_A^P) \) such that \( a(s_A^P) > s_A^P \) and any \( i \in A_Y \) with \( a_i > a(s_A^P) \) chooses \( s_i = a_i \) and any \( i \in A_Y \) with \( a_i < a(s_A^P) \) chooses \( s_i = s_A^P \); and there is a second cutoff \( a(d, s_A^P, s_D^P) \) such that \( a(d, s_A^P, s_D^P) > s_D^P \) and any \( i \in D_Y \) with \( a_i > a(d, s_A^P, s_D^P) \) chooses \( s_i = a_i \) and any \( i \in D_Y \) with \( a_i < a(d, s_A^P, s_D^P) \) chooses \( s_i = s_D^P \). Both of these cutoffs depend crucially on parameter \( K \).

At the first stage, assume in equilibrium \((s_A^P)^* = 1 \) (we later check that \( s_A^P = 1 \) is a best response). Let \( f(d) : \mathbb{R}_+ \rightarrow [0, 1] \) and \( g(d) : \mathbb{R}_+ \rightarrow [0, 1] \) be such that given \( d \), given \( s_D^P = f(d) \), and given equilibrium play in stages 2 and 3, an agent \( i \in D_Y \) with skill \( s_i = f(d) \) is indifferent.
between assimilating or not if all agents in \( D_Y \) with skill level strictly greater than \( f(d) \) assimilate, and an agent with ability \( a_i = g(d) \) is indifferent between choosing skill \( s_i = g(d) \) and assimilating, or choosing \( s_i = f(d) \), again if all agents with higher skill level assimilate. The intuition is that agents with ability \( a_i \in [f(d), g(d)] \) choose skill \( s_i = f(d) \) and do not assimilate, and agents with ability \( a_i \geq g(d) \) choose skill \( s_i = a_i \) and assimilate (note that these functions are not defined for very high values of \( d \), as no agent is indifferent about assimilation in that case). If \( \lambda \) is sufficiently small, \( f(d) \) and \( g(d) \) are strictly increasing. Let the equilibrium \( d^* \) be any \( d \) such that agent \( l \) strictly prefers an agent \( i \in D_Y \) with skill \( s_i = g(d) \) to assimilate, and an agent \( i \in D_Y \) with skill \( f(d) \) to not assimilate. This equilibrium value \( d^* \) leads to assimilation for agents with ability \( a_i \geq g(d^*) \). In equilibrium, \( (s^P_D)^* = f(d^*) \). I check that \( (s^P_D)^* \) and \((d^*, (s^P_A)^*))\) are mutual best responses.

If agent \( m \in D \) deviates to \( s^P_D < (s^P_D)^* \), all agents in \( D \) (including \( m \)) become worse off because those with ability \( a_i \in [s^P_D, a_1(d^*, 1, s^P_D)] \) lower their skill from \( \min\{a_i, (s^P_D)^*\} \) to \( s^P_D \) to avoid the punishment \( K \), leading to a decrease in \( g_D \). If \( k \) deviates to \( s^P_D > (s^P_D)^* \), then agents with ability \( a_i \in [s^P_D, (s^P_A)^*] \) choose \( s_i = a_i \) and assimilate, again reducing \( s_D \). Choosing \( d < d^* \) causes those with skill \( f(d^*) \) to assimilate, which makes \( l \) strictly worse off. Choosing \( d > d^* \) causes those with ability \( g(d^*) \) to not assimilate, which makes \( l \) worse off. Choosing \( (s^P_A)^* < 1 \) again causes those with ability \( g(d^*) \) to not assimilate and it furthers reduces \( s_A \) directly by reducing the average skill of agents with an advantaged background, hence making \( l \) strictly worse off. Therefore, \( (s^P_A)^* = 1 \), \( d^* \) and \( (s^P_D)^* = f(d^*) \) are best responses.

**Part II**: Next I show that \( s^P_D < \gamma \) and \( s^P_A > 1 - \varepsilon \) in all equilibria.

I noted that the third stage does not have a unique solution. If \( \Omega_\varepsilon \) has an interval with density zero, and the cutoff is in this interval, any value in the interval serves as a cutoff. Let \( s_1(d, \Omega) \) and \( s_2(d, \Omega) \) denote the lowest and highest possible values of the cutoff. The solution (and not just the cutoff that represents it) for the third stage may not be unique. Suppose \( s_1(d, \Omega) \in [s^P_D, s^P_D + \varepsilon_1] \) for some small \( \varepsilon_1 \geq 0 \). Suppose that every \( i \in D_Y \) with skill \( s_i \in [s^P_D, s_1(d, \Omega)] \) deviates to \( e_i = 1 \). Then the average skill of each group suffer a discrete change, and if this change increases \( s_A - s_D \), it can make each of the deviators strictly prefer to persist in the deviation. In this case, a second equilibrium arises in which \( s(d, \Omega) = s^P_D - \varepsilon_2 \) for some \( \varepsilon_2 \geq 0 \). Let \( v^{1, J}_i \) and \( v^{2, J}_i \) denote the utilities from externalities derived by agent \( i \) in group \( J \) in the first and second of these two equilibria. However, note that \( \lim_{\lambda \to 0} (v^{2, J}_i - v^{1, J}_i) = 0 \), which implies \( \lim_{\lambda \to 0} \varepsilon_1 + \varepsilon_2 = 0 \). Hence the two stage-game

30
equilibria become arbitrarily close if \( \lambda \) is sufficiently small.

Consider any equilibrium with \( (s^P_A, s^P_D) \) in the first stage. Suppose \( (s^P_D)' \geq \gamma \) for some \( \gamma > \frac{1}{2} \). If \( w_A \) is sufficiently small, then agent \( l \) strictly prefers agents with skill level contained in an open interval around \( s_i = \gamma \) to assimilate. Thus, if \( \lambda \) is sufficiently small so that \( d(s) \) is strictly increasing, it must be that \( d \) is low enough so that agents with skill \( s_i = \gamma \) indeed assimilate (otherwise, agent \( l \) who chooses \( d \) becomes better off deviating to a lower \( d \)). But then, agent \( m \) can become better off deviating to lower \( s^P_D \) to the actual cutoff of assimilation in this equilibrium, so that after the deviation, some agents stop assimilating. Thus \( s^P_D \geq \gamma \) cannot be sustained in equilibrium; in any equilibrium, \( (s^P_D)' < \gamma \).

I also show that \( (s^P_A)' \geq 1 - \varepsilon \). Suppose not. Then agent \( l \) who chooses \( s^P_A \) can deviate to \( s^P_A = 1 \) and simultaneously deviate to a higher \( d \) so that the new cutoff for assimilation is weakly greater than before the deviation. Let \( v^*_i,J \) and \( v'_i,J \) denote the utilities from externalities derived by agent \( i \) in group \( J \) before the deviation, and after the deviation. Note that \( \lim_{\lambda \to 0} (v^*_i,J - v'_i,J) = 0 \), which implies that the distance in the cutoff for assimilation before and after the deviation converges to zero. But then, the fraction of the set of agents \( D_Y \) who change their assimilation decision in response to the deviation converges to zero as \( \lambda \to 0 \), while the fraction of \( A_Y \) who change their skill acquisition decision stays constant at \( 1 - (s^P_A)' \). It follows that for a sufficiently small \( \lambda \), the positive effect of increasing skill of agents with advantaged background and ability in \( [(s^P_A)', 1] \) is greater than the effect due to the change in assimilation decisions. Thus agent \( l \) prefers to deviate, and thus \( s^P_A \leq 1 - \varepsilon \) cannot be sustained in equilibrium, so that \( (s^P_A)' > 1 - \varepsilon \) in any equilibrium, as desired.

References


Bisin, A. and T. Verdier (2001). The Economics of Cultural Transmission and the Dynamics of


