Abstract

I present a theory of assimilation in a heterogeneous society composed of two groups with distinct social norms and unequal background. Members of the group with a relatively disadvantaged background 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 choose to assimilate. The theory provides a novel explanation of the so called “acting white” phenomenon, in which students from disadvantaged ethnic groups punish their co-ethnics who succeed academically. I show that punishing success and thus raising the cost of acquiring skills needed to assimilate is an optimal strategy by low ability students to keep their more able co-ethnics in the disadvantaged group.

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

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

* I am grateful to Drew Conway and Jeff Biddle for their contribution, to Dan Bernhardt, Alberto Bisin, Renee Bowen, Oeindrila Dube, Raquel Fernandez, Paola Giuliano, Stephan Heblich, Loukas Karabarbounis, Rachel Kranton, Maggie Penn, Carlo Prato, Larry Samuelson, Dan Silverman, Jakub Steiner, Will Terry, Leonard Wantchekon, participants at the 2012 NBER Political Economy Summer Institute and audiences at talks at MPSA 2010, LSE, Northwestern-Kellogg, Stanford-GSB, Chicago, Harvard/MIT, NYU, Oslo, USC, UPF, Bristol, Leicester, MSU and ITAM 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. I am also affiliated to the Department of Economics at the University of Bristol, and I greatly benefited from its support.

† Email: eguia@msu.edu. Mail: 486 West Circle Dr, 110 Marshall-Adams Hall, Dept. Economics, Michigan State U. East Lansing, MI 48824.

“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, and those with a disadvantaged background. Agents are characterized by their background and their ability. Disadvantaged agents choose whether or not to assimilate by joining the advantaged group. Agents have an effect on the group to which they ultimately belong: more skilled agents have a more positive effect. 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 improve the group. 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.

This 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).

I explain why 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 benefit 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.

This explanation fits the empirical findings of Fryer and Torelli (2010) on acting white and explains why African-American and Hispanic students, but not white students, experience a negative correlation between popularity and high grades.

The traditional explanation of acting white (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. McWhorter (2000) argues that African-Americans engage in self-sabotage: 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 are distinguished by their social type (low or high) and their economic type. Students with low social type are socially inept: other agents derive more utility from ostracizing them than from socializing with them. In equilibrium, socially inept agents choose to study and are shunned by their peers, while other students differentiate themselves from the socially inept by choosing not to study. While compelling, this reasoning applies to all ethnicities, and thus it cannot explain the asymmetry across ethnic groups which is the essence of the acting white phenomenon.

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

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

This paper builds upon an extensive literature on theories of social identity formation.² The literature on the economics of culture argues that minorities adopt and pass on to their descendents identities that are anti-achievement (Akerlof and Kranton 2000), traditional (Bénabou and Jean Tirole 2011) or ethnic (Bisin and Verdier 2000 and 2001; Carvalho and Koyama 2013) because if they shed this identity and embrace the productive/modern/majority identity, they suffer an exogenously given cost. Shayo (2009) and Klor and Shayo (2010) theorize that agents would like to identify with a high status group formed by agents similar to them.

Identity theories teach us that given a sufficiently high exogenous cost of assimilation, it is not optimal to assimilate. I propose a theory that recognizes that the difficulty of assimilation is endogenous: it depends on the actions of the agents with an advantaged background. The opportunities for friendship and social connections, and the nature of an agent’s social interactions depend less on her own identity (her concept of self) and more on how she behaves, on what other agents think of her, and on how they treat her as a result. Identity theories do not ask why agents with an advantaged background discriminate

---

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

²For interdisciplinary perspectives on identity, see the surveys by Hogg (2003) in social psychology; Hill (2007) in law and economics, and Jenkins (1996) in all the social sciences.
against those who seek to assimilate: I show that discrimination arises in equilibrium as agents pursue their own selfish interests.

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

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

1 The Model

Consider a society with a continuum of agents of unit mass. Agents are distinguished by their background and their ability, both of which are exogenously given. The background of a half of the agents is advantaged. Let \(A\) denote the set of agents with an advantaged background. The other half of the agents, denoted by \(D\), have a disadvantaged background.\(^3\)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 (Fearon and Laitin 2000; Caselli and Coleman 2013) or urban violence (Silverman 2004) face a different strategic environment.
Let $\theta_i$ denote the exogenously given ability or talent of agent $i$. Individual ability is private information. Assume that for each set of agents $\mathcal{J} \in \{A,D\}$ the distribution of ability over $\mathcal{J}$ is uniform in $[0,1]$.

Agents choose their skill and their social group.

**Choice of skill**

Let $s_i$ be the skill of agent $i$. Skill is endogenous, strategically chosen by agent $i$, subject to constraints given by their background and ability. An agent’s innate ability is an upper bound on how skilled the agent can become. For an agent $i$ with an advantaged background, her innate ability is the only constraint, and she chooses skill $s_i \in [0,\theta_i]$. Agents with a disadvantaged background face some exogenously given handicap (deficient schools, less-educated parents, poverty, etc.) that limit their acquisition of skills. An agent $i \in D$ chooses skill $s_i \in [0,g\theta_i]$, where $g \in \left(\frac{1}{2},1\right)$ is a productivity parameter such that $1 - g$ captures the inequality of opportunity to acquire skills affecting the agents with a disadvantaged background ($g = 1$ would represent perfect equality of opportunity). Let $s$ be the skill choice of every agent.

I compare two games: A benchmark game without peer punishments, and a game with peer punishments in which each agent $i$ is susceptible to peer pressure by other agents from her own background—who attend their same school. I model peer punishments as follows: for each background $\mathcal{J} \in \{A,D\}$, let $l_{\mathcal{J}} \in \mathcal{J}$ be an agent with ability $\theta_{l_{\mathcal{J}}} \leq \frac{1}{2}$. Assume $l_{\mathcal{J}}$ chooses a skill threshold $s_{l_{\mathcal{J}}}^P \in [0,1]$, which is observed only by every $i \in \mathcal{J}$. Assume that every $i \in \mathcal{J}$ who chooses $s_i > s_{l_{\mathcal{J}}}^P$ incurs a fixed cost $K \geq 0$. This $K$ 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. The case $K = 0$ captures the game without peer punishments, and $K > 0$ the game with peer punishments.

**Social groups**

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

Notice that I use calligraphic letters $\mathcal{J} \in \{A, D\}$ 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.

**The cost of assimilation**

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

The cost of assimilating is $d(1-s_i)$, where $d \geq 0$ is the difficulty of learning and embracing the patterns of behavior consistent with membership in $A$. I assume that the cost function is decreasing in $s_i$ to capture the intuition that to adapt is easier for more skilled agents. 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. Initially, assume an arbitrary agent $h \in A$ chooses $d$.

**Utility function**

Agents derive utility from their own skill, and the skill of those in her same social group.
Agents are positively affected by the skill of those in their group because 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, and more highly skilled acquaintances provide better opportunities and more positive interactions.\footnote{For recent experimental evidence on the economic benefits of social interaction, see Feigenberg, Field and Pande (2013).} Because interactions among agents are more fruitful if both agents share a common skill, an agent’s own skill and the skill of those in her group are complementary, in the sense that having more of one increases the marginal utility of having more of the other.

Formally, for any $J \in \{A, D\}$, let $s_J$ be the average skill of agents in $J$. Assume that an agent with skill $s_i$ in social group $J \in \{A, D\}$ with average skill $s_J$ derives a utility $s_i s_J$, and in addition she may experience costs of assimilation or peer punishments.

Let $U_i(d, s_J^P, s, a)$ denote the utility function of agent $i$ as a function of the discrimination level $d$, the punishment threshold for agents of her background, the skill of every agent, and the assimilation decisions of all agents in $\mathcal{D}$. If we let $a_i$ be exogenously fixed at 0 for any $i \in \mathcal{A}$, and we let $1_{s_i > s_J^P}$ be the indicator function such that $1_{(s_J^P, 1)}(s_i) = 1$ if $s_i > s_J^P$ and $1_{(s_J^P, 1)}(s_i) = 0$ otherwise, then the utility of an agent $i$ with background $J \in \{A, D\}$ in social group $J \in \{A, D\}$ can be written as:

$$U_i(d, s_J^P, s, a) = s_i s_J - 1_{(s_J^P, 1)}(s_i) K - a_i d (1 - s_i) \quad (1)$$

Every agent derives utility from her own skill and the average skill of the social group they join, and incurs a disutility $K$ if their skill is above the punishment threshold for their background; further, agents with a disadvantaged background who assimilate ($i \in \mathcal{D}$ such that $a_i = 1$) incur the cost of assimilation $d(1 - s_i)$.

**Timing**

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

1. Agent $h \in \mathcal{A}$ chooses the difficulty of assimilation $d$ and, simultaneously, for each $J \in \{A, D\}$ agent $l_J$ chooses a peer pressure threshold $s_J^P$. Every $i \in \mathcal{D}$ observes $d$ and $s_J^P$. 
but not $s^P_A$, and every $i \in A$ observes $s^P_A$.\footnote{Whether agents with an advantage background observe $s^P_J$ or $d$ is irrelevant for the theory.}

2. Each agent $i$ chooses her skill $s_i$. This choice is private information.\footnote{Peer punishments are blackboxed for expositional simplicity. If we wished to model agent $I_J$ meting out the punishment to any agent $i \in J$ who chooses $s_i > s^P_J$, then we would assume that $i$ and $I_J$ observe $s_i$. The relevant assumption is that agents in $A$ do not observe $s_i$ and cannot choose skill-contingent discrimination, which would make screening for high-skilled types trivial.}

3. Agents in $D$ choose whether to assimilate or not. Payoffs accrue.

## 2 Results

As a preliminary observation, note that the first best outcome for the agents with an advantaged background consists on screening agents with a disadvantaged background, so that only those with a sufficiently high skill level assimilate, in such a way that it maximizes the average skill level of those who end up belonging to the advantaged group (those born into it, and those assimilated into it).

**Lemma 1** The first best outcome for agents with an advantaged background is such that each of them chooses maximal skills according to her potential ($s_i = \theta_i$ for any $i \in A$); agents with a disadvantaged background and ability above cutoff $\theta^* = 2 - \sqrt{3 - \frac{1}{g}}$ choose maximal skills and assimilate ($s_i = g\theta_i$ and $a_i = 1$ for any $i \in D$ such that $\theta_i > \theta^*$); and agents with a disadvantaged background and ability below cutoff $\theta^*$ do not assimilate.

I first show that in the game without peer punishments, agents with an advantaged background are able to attain this first best (for them) as an equilibrium outcome by choosing a particular level of discrimination.

I solve by backward induction, finding perfect Bayesian equilibria.

Given $d$, and given any strategy profile by all other members of $D$, an agent $i \in D$ prefers to assimilate only if her skill $s_i$ is high enough so that her cost of assimilating $d(1 - s_i)$ is sufficiently small. It follows that for any $d$, there is a cutoff correspondence $\sigma(d)$ from the discrimination level to the level of skill such agents in $D$ pick a cutoff $\hat{\sigma} \in \sigma(d)$ and as a joint best response each of them assimilates if and only if her skill is above $\hat{\sigma}$. 
The cutoff that is most favorable for advantaged agents is

\[ \hat{\sigma}^* = \arg \max_{\{x\}} s_A(x) \text{ s.t. } s_A(x) = \frac{g + (g - x)(g + x)}{2(2g - x)}, \]

where \( s_A(x) \) is the average skill of the agents in \( A \) as a function of \( x \) given that agents in \( D \) assimilate if and only if their skill is above \( x \). I show in the appendix that there is a unique \( d^* \) such that if agent \( h \in A \) chooses \( d^* \), the cutoff for assimilation becomes \( \hat{\sigma}^* \). In this unique equilibrium, agents with an advantaged background attain their first best: the most able and most skilled of the agents with a disadvantaged background assimilate. Formally:

**Proposition 2** Assume \( K = 0 \). Given any parameter \( g \in \left( \frac{1}{2}, 1 \right) \), there is a unique Perfect Bayesian equilibrium and this unique equilibrium delivers the outcome that is first best for agents with an advantaged background.

The intuition is that agents with an advantaged background are able to optimally screen those who assimilate and join their group, by setting a positive but not too large difficulty of assimilation so that only agents with high ability (who in equilibrium are highly skilled) assimilate. As noted in Lemma 1, the first best assimilation for the agents with an advantaged background sets a cutoff at exactly the skill level that maximizes the average skills \( s_A(x) \) in the resulting advantaged group after assimilation (bold solid curve in Figure 1). The benefit of assimilating for a given agent \( i \in D \) with skill \( s_i = x \) (thin solid line in Figure 1) is equal to the difference \( x(s_A(x) - s_D(x)) \), which is a strictly concave function of the assimilation cutoff \( x \); whereas, the cost function is linear with slope \(-d\) (dashed line in Figure 1) and thus it is possible to fix \( d = d^* \) such that the cutoff occurs at exactly the skill level \( x^* \) that maximizes the average skill level of the advantaged group. This is the first-best for agents with an advantaged background: the difficulty of assimilation is low enough so that every agent whose assimilation benefits the advantaged group chooses to assimilate, while at the same time it is high enough that agents whose assimilation would be detrimental to the advantaged group choose not to assimilate.\(^7\)

\(^7\)I calculate this exact goldilocks level \( d^* \) as a function of the inequality parameter \( g \) as part of the proof of Proposition 2 in the Appendix.
Now consider a game with peer punishments, i.e. $K > 0$. I solve by backward induction. First I explain the intuition, then I state the result.

At step 2, any agent $i \in A$ chooses skill $s_i \in \{s^P_A, \theta_i\}$ and any $i \in D$ chooses $s_i \in \{s^P_D, g\theta_i\}$. At step 1, agent $l_A \in A$ has no incentive to punish any agent with her background, because given the complementarity between own individual skill and own group's average skill, a higher skill level for any $i \in A$ generates a higher utility to all members of $A$. Hence, in equilibrium, $s^P_A = 1$.

Whereas, agent $l_D \in D$ who chooses $s^P_D$ 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 $A$ and $D$. For any $\Omega$, there is a threshold function increasing in $d$ such that in the equilibrium of a reduced assimilation game played given a fixed pair $(d, \Omega)$, 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 $s^P_D$ below the threshold of assimilation deters some agents in $D$ from acquiring a skill level above the threshold and thus from assimilating. The optimal peer pressure maximizes $s_D$ by inducing as many highly able agents as possible to stay in the disadvantaged group $D$, while lowering their skill level.
only just as much as it is necessary to prevent them from assimilating. Hence in every
equilibrium, \( s^P_D < g \) and some agents with a disadvantaged background are deterred from
overachieving.

The following proposition shows that in equilibrium, highly able disadvantaged students
are punished by their peers if they acquire a high level of skills (any level between \((s^P_D)^* \) and
\( g \)), while students with an advantaged background are never punished for acquiring skills.
Further, I identify one equilibrium in which agents with a disadvantaged background who
acquire skills beyond the average skill level \( s_A \) of advantaged agents are punished by their
peers.

**Proposition 3** Assume \( K > 0 \). For any \( g \in \left(\frac{1}{2}, 1\right) \), in any Perfect Bayesian equilibrium,
peer punishment thresholds are \((s^P_D)^* < g \) and \((s^P_A)^* = 1 \). Furthermore, there exists an
equilibrium in which \((s^P_D)^* = s_A \) and \((s^P_A)^* = 1 \).

Proposition 2 showed that the equilibrium without peer pressure leads to assimilation.
This assimilation of their better co-ethnics makes the remaining (lower-skilled) agents with
a disadvantaged background worse off. 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.

Agents with a disadvantaged background and low ability, who are harmed by the assim-
ilation 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, which harms agents with low ability
and a disadvantaged background. In every equilibrium, highly able agents with a disad-
vantaged background are pressured to underperform; whereas, agents with an advantaged
background are not. This is the acting white phenomenon.

I illustrate this phenomenon with a numerical example.

**Example 1** Let \( g = 0.6 \). Let \( U_A, U_{D-} \) and \( U_{D+} \) respectively denote the average utility of
\( \{i \in A\}, \{i \in D\} \) and \( \{i \in D : \theta_i \leq \frac{1}{2}\} \). Columns 2 and 3 in the table below compare the
equilibrium outcomes under an assumption of no peer pressure \((K = 0)\) in column 2, and peer pressure \((K = 0.03)\) in column 3, where \(s^P_A = 1\) and \(s^P_D = s_A\) as part of the equilibrium.

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2) Full Segregation</th>
<th>(3) No Peer Pressure</th>
<th>(4) Peer Pressure</th>
<th>(4)-(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d^*)</td>
<td>n.a.</td>
<td>0.2610</td>
<td>0.2401</td>
<td>-0.0209</td>
</tr>
<tr>
<td>(\theta^*)</td>
<td>n.a.</td>
<td>0.8453</td>
<td>0.9104</td>
<td>+0.0651</td>
</tr>
<tr>
<td>(s^D)</td>
<td>0.3</td>
<td>0.2536</td>
<td>0.2716</td>
<td>+0.0180</td>
</tr>
<tr>
<td>(s^A)</td>
<td>0.5</td>
<td>0.5072</td>
<td>0.5060</td>
<td>-0.0012</td>
</tr>
<tr>
<td>(U^A)</td>
<td>0.25</td>
<td>0.2536</td>
<td>0.2530</td>
<td>-0.0006</td>
</tr>
<tr>
<td>(U^D)</td>
<td>0.09</td>
<td>0.0798</td>
<td>0.0813</td>
<td>+0.0015</td>
</tr>
<tr>
<td>(U^D-)</td>
<td>0.045</td>
<td>0.0380</td>
<td>0.0407</td>
<td>+0.0027</td>
</tr>
</tbody>
</table>

While the acting white equilibrium makes agents with an advantaged background worse off than the assimilation equilibrium with no peer pressure, it increases the average skill level in the disadvantaged social group and, crucially, it makes agents with low ability and a disadvantaged background –the perpetrators of peer punishments– strictly better off.

Figure 2 summarizes the effects of the acting white phenomenon on agents with a disadvantaged background. The horizontal axis measures ability. Students with ability below cutoff \(\frac{1}{g}s^P_D\) are not subject to any peer pressure, because \(\theta_i \leq \frac{1}{g}s^P_D\) together with \(s_i \in [0, g\theta_i]\) implies \(s_i \leq s^P_D\): their ability constraints them to acquire skills below the punishment cutoff \(s^P_D\). Students with ability above \(\frac{1}{g}s^P_D\) are subjected to peer pressure to underperform (to acquire skills below their potential). Those with ability between \(\frac{1}{g}s^P_D\) and the equilibrium assimilation cutoff \(\theta^*\) yield to this pressure and underperform (choosing \(s_i = s^P_D < g\theta_i\)) to
escape social punishments. Whereas, the most able reject the peer pressure, endure the consequent alienation from their co-ethnics, and ultimately assimilate into the advantaged community.

3 Empirical Implications

This theory of assimilation and discrimination yields several testable predictions, and other insights that help explain known stylized facts.

First, with respect to immigration, the theory predicts that high-skilled immigrants assimilate, and low-skilled ones do not. The following prediction follows directly from Proposition 2:

**Testable Hypothesis 1:** *Higher-skilled immigrants from a poor ethnicity are more likely to assimilate than lower-skilled immigrants from that same ethnicity.*

This hypothesis is testable, using data on education level as proxy for skills, and out-group marriages (intermarriage outside the ethnicity) as proxy for assimilation.\(^8\)

Descriptive statistics for the subset of Hispanic-American immigrants from the 1980 and 2000 US Census support this hypothesis. Uniquely among cultural identities, the Census asks subjects to identify as either Hispanic or non-Hispanic, separately from any racial category. I code an out-group marriage to a non-Hispanic as assimilation \((a_i = 1)\) and an in-group marriage to another Hispanic as non-assimilation \((a_i = 0)\). Consider the subsample of Hispanics who immigrated from Mexico, Central, or South America unmarried and at age older than 18 (so that their assimilation choices once settled in the US did not affect their education outcomes), and who were 23 years of age or older and married in 1980, and divide them into four categories according to their education (less than high-school diploma, high-school diploma, some college, 4+ years of college), we find the following marriage patterns.

The first row indicates the percentage of out-group marriages to non-Hispanics among males, and the third among females. The pattern is qualitatively very similar for both

\(^8\)This prediction only applies to immigrants of ethnicities that are poorer than the general population of the host country (for instance Hispanics in the United States) and not to immigrants whose ethnic population has higher average skills than the host country at large (for instance, Asian-Americans).
Table 1: Outgroup marriages by Hispanic immigrants, by education level.

<table>
<thead>
<tr>
<th>% outgroup marriages (men)</th>
<th>&lt;HS</th>
<th>HS</th>
<th>Some college</th>
<th>4+ y. college</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subsample size (men)</td>
<td>4.6%</td>
<td>12.1%</td>
<td>20.6%</td>
<td>30.9%</td>
<td>9.1%</td>
</tr>
<tr>
<td>% outgroup marriages (women)</td>
<td>6.9%</td>
<td>26.8%</td>
<td>39.6%</td>
<td>55.3%</td>
<td>17.2%</td>
</tr>
<tr>
<td>Subsample size (women)</td>
<td>1933</td>
<td>648</td>
<td>336</td>
<td>159</td>
<td>3076</td>
</tr>
</tbody>
</table>

genders: among Hispanic who immigrated as unmarried adults and later married, those with 4 years of college are about four times as likely to marry non-Hispanics as those with less than four years of college or no college education at all (30.9% to 7.7% for males; 55.3% to 15.1% for females).\(^9\) This table reveals a correlation between skill acquisition and outgroup marriages; it does not establish a causal link. Note that whether a causal link exists or not, a positive correlation between skills and assimilation suffices to establish that the most skilled immigrants (tend to) assimilate and enrich the advantaged social group, whilst the less skilled ones stay out, as in Proposition 2.\(^10\)

This empirical finding not only supports the theory, but it also helps to rule out an alternative hypothesis that could have cast a doubt on the theory: if the scarcity of highly skilled members in the disadvantaged ethnic group allows them to attain positions of leadership or other rewards within their group, then high-type agents have an incentive to stay in this group, and it is conceivable that this counter-incentive could overwhelm the incentives based on skill complementarity, so that within the disadvantaged group, high skilled agents would more likely to stay in the group and low skilled agents more likely to leave and assimilate. This alternative view is refuted by the evidence.

Consider as well a comparative static result on the effect of inequality on assimilation. As \(g\) increases and conditions improve in the disadvantaged group, more agents with a disadvan-

---

9I use the 1980 Census because it is the only census that asked year of marriage. To replicate the analysis with data from the 2000 Census (the last one with a comparably large sample) and exclude agents who were already married at the time of immigration I include only migrants who immigrated at age 11 through 18. The percentage of outgroup marriages to non-Hispanics increases an order of magnitude from approximately 2% for both males and females with less than high school education, to 25% and 36.8% for males and females with at least 4 years in college (a full table is available from the author).

10On the other hand, we need to identify a causal link to establish that the peer pressure equilibrium in Proposition 3 deters exit from the disadvantaged group as predicted. I am not aware of any empirical result confirming or refuting the existence of this causal link.
taged background assimilate, and the average skill level increases as well in the advantaged group. That is, an improvement in the exogenous productivity parameter of the disadvantaged group leads more agents to leave this group and makes agents with an advantaged background (who are assumed to experience no productivity shock) strictly better off in equilibrium as a result of this increased assimilation.

**Corollary 4** In the unique equilibrium, the fraction of agents with a disadvantaged background who assimilate and the average skill of both the advantaged and disadvantaged groups, are strictly increasing in $g$ for any $g \in \left(\frac{1}{2}, 1\right)$.

This follows almost directly from the proof of Proposition 2: as agents with a disadvantaged background become in distribution more skilled, more of them are more skilled than the average member of the advantaged group, so more of them are incentivized to assimilate, raising the average skill level in the advantaged group.\(^{11}\)

**Testable Hypothesis 2:** Greater economic inequality across ethnic groups leads to less assimilation.

This empirical implication can be tested using data on inequality across ethnic groups as an independent variable, and on intermarriages as the dependent variable.

The theory is also consistent with empirical findings on the attitudes of natives toward immigration. Using data from the United States, Hainmueller and Hiscox (2010) find that all native agents alike, poor or rich, high-skilled or low-skilled, prefer high-skilled immigration to low-skilled immigration, and in fact oppose the latter. This finding goes against the labor market competition explanation,\(^{12}\) which led Hainmueller and Hiscox (2010) to argue rather forcefully that: “economic self-interest, at least as currently theorized, does not explain voter attitudes toward immigration.” From their abstract:

\(^{11}\) The equilibrium average skill of the disadvantaged group is proportional to the equilibrium average skill of the advantaged group, and thus it increases as well. Through the assimilation process, the advantaged group extracts some (but not all) of the gains that come to agents with a disadvantaged background from an increase in $g$.

\(^{12}\) Labor market competition theories such as Becker’s (1957) predict that natives oppose entry by immigrants with similar skills, but favor entry by immigrants with very different skills. If so, high-skilled natives would oppose high-skill immigration and favor low-skill immigration.
“The labor market competition model predicts that natives will be most opposed to immigrants who have skill levels similar to their own. We find instead that both low-skilled and highly skilled natives strongly prefer highly skilled immigrants over low-skilled immigrants, and this preference is not decreasing in natives’ skill levels. The fiscal burden model anticipates that rich natives oppose low-skilled immigration more than poor natives [...]. We find instead that rich and poor natives are equally opposed to low-skilled immigration.”

The theory in this paper provides an explanation of these attitudes that is based strictly on self-interest that fully accounts for these attitudes: immigrants who are so skilled that they exert positive externalities are welcome, while others are not.\textsuperscript{13} The theory is also consistent with immigration policies that offer a path to naturalization and assimilation for highly skilled immigrants (such as the “green card” in the U.S. or the “blue card” in the E.U.), while they keep the bulk of low skilled immigrants as undocumented or temporary guest worker aliens. As theorized, highly skilled immigrants are embraced as valuable members of society, whereas low skilled immigrants are not welcome to participate in civil society even when their labor is used as a production factor in the economy.

In the next subsection I discuss how the theory explains and conforms to known stylized facts about the acting white phenomenon. Additional empirical implications follow from extensions to the theory discussed in the following section.

3.1 Acting White

“Acting white” is “\textit{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 “\textit{the popularity of white students increases as their grades increase}. For black and Hispanic students, there is a drop-off in popularity for those

\textsuperscript{13}Aside from labor market competition, Heinmueller and Hiscox also consider and critique theories based on the cost of providing public services (Hanson, Scheve and Slaughter 2007). However, other theories of economic self-interest (see Storesletten 2000) can also explain their findings based on the net positive contribution of high-skilled immigrants to fiscal balances.
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.

Proposition 3 provides a game-theoretic explanation of the acting white phenomenon: students in under-privileged communities dissuade their co-ethnics from acquiring skills in order to increase the cost of assimilation and deter exit from the community. This explanation has distinct empirical implications from those of alternative explanations in the literature (see the survey by Sohn 2011).

The “oppositional culture” theory of Fordham and Ogbu (1986) and Fordham (1996) posits that academic failure is an integral part of African-American group identity: while whites embrace studiousness and hard work, minorities reject these values, embracing instead a counterculture defined in opposition to 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.” They trace back the roots of black students’ self-identification with academic failure to a history of oppression in which whites negated their accomplishments.

Even if correct at the time, this account is anachronistic: 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.  

A second now traditional explanation is the “self-sabotage” argument: African-Americans engage in willful victimism, persuading themselves that discrimination in the job market makes costly accumulation of human capital not worthwhile (McWhorter 2000). To the extent that self-saboteurs are deemed unworthy of social assistance, the term “self-sabotage”

---

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

15 Silverman (2004) provides clear micro-foundations for a different kind of oppositional culture: a “street culture” that induces agents to act violently in order to gain a reputation for toughness. While acting peacefully in Silverman’s streets leads to costs similar to those of acting white, a key feature of the acting white phenomenon is that African-Americans want other African-Americans to not act white; whereas, in Silverman’s theory of urban conflict every agent wants all other agents to act peacefully.
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 Survey 2010).

The oppositional culture and the sabotage theories imply that the acting white problem ought to be more severe in schools that are more segregated, and thus not exposed to alternative cultures or views. The screening theory in this paper has the opposite empirical implication: the explanation of acting white as a strategic exit barrier only applies where an alternative to the disadvantaged culture exists and exit to a different social group is an option, such as in less segregated (not all-black) schools. In a fully segregated environment that has no little or no contact with outside groups, in which assimilation appears infeasible, the screening theory does not apply and it does not predict the emergence of acting white.

**Testable Hypothesis 3:** The acting white phenomenon occurs in ethnically diverse schools, and not in homogeneously black ones.

Miron and Lauria (1998), Tyson (2006), Fryer (2006) and Fryer and Torelli (2010) test something close to this hypothesis. They all find that the acting white problem is more severe in less segregated (that is, in more racially integrated) schools: in predominantly black schools, where assimilation would appear least feasible, “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: black students are exposed to interaction with white students only in mixed schools, so students in these schools have greater opportunities to join a predominantly white social network and abandon the black community. A top student in a fully segregated school cannot shun the black community, as there is no alternative community to join, so acting white does not occur.
Fryer (2006) conjectures that perhaps the problem is attenuated if school desegregation leads to cross-ethnic friendships. The screening theory implies 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.

Fryer and Torelli (2005) note that the oppositional culture and the sabotage theories “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. They model a set of students and a black-boxed “peer group.” Each student’s type has two components: an “economic type” and a binary (high or low) “social type.” The economic type determines the marginal utility of time spent studying. The social type determines the marginal utility of leisure time, and it also determines the utility that the peer group derives if it accepts a student (positive if the social type of the accepted student is high, but negative if it is low). Students allocate a unit of time between studying and leisure. In equilibrium, those with a low social type, who face a lower opportunity cost of studying because their marginal utility from leisure is low, study more. The peer group shuns those who study, and accepts those who signal their high social type by not studying. Some high social types thus choose not to study.

While this argument is compelling, it should apply to all races and social groups: it 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.

This unexplained 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 a parameter that measures the quality of schools in disadvantaged areas. Solely from this unequal technology factor, it follows that agents with a disadvantaged background discourage their peers’ acquisition of skills, while agents with an advantaged background
never discourage skill acquisition.

The signaling theory by Austen-Smith and Fryer (2005) and the screening theory in this paper disagree on 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, according to the signalling argument, the popularity of African-American and Hispanic students among students of other ethnicities must decrease. According to my screening theory, 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). So here we have a test to discriminate among the two theories.

**Testable Hypothesis 4:** The relation between the grades of African-American (and Hispanic) students and their popularity among white students is strictly positive.

Austen-Smith and Fryer (2005)'s theory leads to the opposite testable empirical implication: a strictly negative relation between grades and number of friends.

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.” Pattacchini and Zenou (2015) find that higher test scores are associated with a higher percentage of out-race friends for black teenagers, that is, the black students with white friends are predominantly the ones with high grades. To find African-American (and Hispanic) students with high grades have white friends, and African-American (and Hispanic) students with lower grades have co-ethnic friends is fully consistent with the screening theory.

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

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

**Testable Hypothesis 5:** *The probability that African-American (and Hispanics) adults remain in the social network of their childhood decreases in educational attainment.*

Remaining in the same network can be measured by geography (residence in the same zipcode), or by friendships, measuring the overlap between the list of closest friends in childhood/adolescence and in adulthood.

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

---

16Hypothesis 5 is the converse of the motivation for the *Moving to Opportunity (MTO)* housing program: MTO posits that if children from a poor background move to a better neighborhood, they do better in later life (see most recently Chetty, Hendren and Katz 2015); I posit if they do better to begin with, they later move to a better neighborhood.
in Britain, high-achievers have suffered a negative externality from their peer group (see Fryer 2007a or Sohn 2011 for a discussion). Hoff and Sen (2006) report a strikingly similar problem in the context of informal insurance provided by extended families in the developing world: “If the kin group foresees that it will lose some of its most productive members as the economy opens up, it may take collective actions ex ante to erect exit barriers.” Jensen and Miller (2011) find evidence of this phenomenon in rural India: parents strategically underinvest in education for their children to prevent them from migrating to the cities. I interpret the acting white phenomenon as yet another exit barrier.\footnote{Religious doctrines opposing secular education or inter-faith marriage can also be understood as exit-deterrence strategies. See, for instance, Carvalho and Koyama (2011).}

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, we can test if an “acting urban” phenomenon, with no racial component, also arises.

**Testable Hypothesis 6:** Rural students who obtain top grades are less popular, regardless of their race, than their classmates with lower grades.

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.”

### 4 Extensions

I consider several extensions and generalizations to the theory.

a) **Generalizations of the functional forms.**

The linear form of the utility from skill and of the cost function capture the intuition of the screening mechanism in a simpler model, but we can consider a utility function with a general functional form \( \varphi(s_i, s_J) \), with the properties that this function be strictly increasing...
in skill $s_i$ and in the average group skill $s_j$, and that the cross-partial derivative own skill and average group skill be non-decreasing \( \left( \frac{\partial \phi(s_i, s_j)}{\partial s_is_j} \geq 0 \right) \).

We can also introduce a general functional form for the cost of assimilation $c(d, s_i)$. For any arbitrary functional form $\varphi(s_i, s_j)$ that satisfies these minimal conditions, the cost function needs to be sufficiently decreasing in own skill $s_i$ to guarantee equilibrium uniqueness given the form of the utilities from skill. Specifically, the derivative of the cost must be sufficiently negative relative to the magnitude of the cost, so that the cost faced by a more skilled agent is only a small fraction of the cost paid by a less skilled agent. If so, agents with unequal ability face such different incentives, the equilibrium is unique separating agents with skill above and below the cutoff. An earlier version of theory with these generalizations and other extensions is available from the author.\(^{18}\)

b) Majority-minority and multiple groups

The theory and all results are robust if set $A$ is a *majority* of size $\lambda > \frac{1}{2}$ and set $D$ is a *minority* of size $1 - \lambda$. I assumed $\lambda = \frac{1}{2}$ to present the intuition in the simplest model. With unequal sizes, the intermediate expressions become more cumbersome, but the main results are unchanged. As the size of the majority increases, its total benefit from the assimilation process is reduced in a transparent way: the positive effect of assimilating high-skilled agents from the minority is less pronounced if the minority is smaller. The cutoff in Lemma 1 is lower; Propositions 2 and 3 hold as stated (the generalized proofs are available from the author).

Identical results hold in a more general model with a majority group and $k$ minority group with group-specific inequality parameters $g_1, \ldots, g_k$, if we assume that assimilation only happens to the majority, and we let the majority discriminate differently against each minority, choosing $d_k$ as a function of $g_k$. We could also envision a model with fully endogenous group choice across multiple groups; I conjecture that in a general equilibrium, some groups of intermediate standing experience out-migration toward an elite group, and in-migration

\(^{18}\)These extensions include a symmetric, sorting model in which each group is advantaged on a different dimension, and assimilation and discrimination occurs in both from each group to the other symmetrically, as agents with heterogeneous preferences seek to join the social group that is advantaged in the dimension they most care about. This material, together with anything reported as “available from the author” is accessible here: http://dl.dropbox.com/u/9574908/OnlineAppDnAJuly2015.pdf
by relatively high-skilled agents from a worse-off group who are shut out by the elite group.

c) Discrimination based on attributes uncorrelated with skill

Notice that if the cost of assimilation were not decreasing, but constant in skill, the uniqueness result would fail: the equilibrium would hold, but another equilibrium would emerge in which no agent assimilates. A flat assimilation cost has a clear substantive interpretation: it captures what we may call “indiscriminate” (non-selective) discrimination borne equally by every agent who assimilates, regardless of her skills. For instance, discrimination based on skin color or other immutable attributes. This type of uniform discrimination based on immutable attributes is qualitatively different from the more “discriminating” (selective) discrimination that hits low-skilled agents harder and touches higher skilled agents more lightly, such as discrimination based on speech patterns or language proficiency. Discrimination based on malleable traits, which high-skilled agents can change at a lower cost, allows for effective screening of those who assimilate by making it possible for high-skilled agents to adopt the traits that escape discrimination, while low-skilled agents find this option too costly and do not assimilate.\footnote{Discrimination based on malleable traits allows advantaged agents to screen high skilled from low skilled agents, without a need to observe actual skill levels. Theories of statistical discrimination show that the inability to observe individual skill causes firms (Fang 2001; Moro and Norman 2004) or a social planner (Norman 2003) to misallocate high-skilled agents to unskilled jobs. In contrast, in this manuscript’s theory, agents with a disadvantaged background sort themselves and since each agent knows her own skill, in equilibrium no agent is misallocated.} Uniform discrimination based on immutable attributes is not a good screening device, and it cannot be rationalized as stemming from an instrumental motivation. If agents with an advantaged background are instrumentally motivated and they can choose whether to discriminate on traits that correlate with skill or on attributes that are not correlated with skill, they must choose to discriminate on traits that correlate with skill.

d) Homophily

The instrumental value of screening for high-skill is not a compelling explanation of discrimination based on immutable attributes (notably skin color). In-group preferences (homophily) provide a simpler explanation: if agents with an advantaged background have an intrinsic preference for this background, and they experience a sufficiently large direct
disutility from the presence of agents with a disadvantaged background in their social group, then they prefer none of them to assimilate, than to screen for high skills.

Consider a more general extension in which agents with an advantaged background have in-group preferences captured by an homophily parameter $\gamma \in \mathbb{R}_+$ so that their utility decreases proportionally in the number of agents with a disadvantaged background who assimilate. Dropping the terms on the cost of peer pressure and cost of assimilation that do not apply to agents with an advantaged background, assume that the utility function of each $i \in A$ is:

$$U_i(s, a) = s_i s_A - \gamma \int_0^{1/\gamma} a_i d\theta_i. \quad (2)$$

Recall that given that agents with a disadvantaged background assimilate if and only if their skill is above a cutoff $x$, the average skill level in the advantaged group $s_A(x)$ is a function of this cutoff that is maximized where $s_A'(x) = 0$. Note that with in-group preferences and such a cutoff for assimilation, the homophily term in the utility function is $-\gamma(1 - x/\gamma)$ and thus, assuming $\gamma$ is small enough to guarantee an interior solution, now the utility of the advantaged group is maximized where $s_A'(x) + \gamma/\gamma = 0$, which is at a higher level of the cutoff $x$. Therefore, fewer agents assimilate in a society in which those with an advantaged background have in-group preferences.

**e) Collective choice of discrimination**

With the benchmark utility function given by expression (1) and no in-group preferences, the identity of the advantaged agent $h$ who chooses the discrimination level $d$ is irrelevant: all agents with an advantaged background have identical preferences to maximize $s_A$, so $h$ can be interpreted as a representative advantaged agent. With in-group preferences given by utility function (2), agents with an advantaged background have conflicting interests about assimilation: they all want agents with skills above a cutoff to assimilate, but they diverge in their preferred cutoff.

Due to the complementarity between own skill and group skill, high-skilled agents benefit more from the increase in average skill that results once other high-skilled agents assimilate. If the homophily parameter $\gamma$ is common across agents, it follows that high-skilled advan-
taged agents gain more on net utility terms than low-skilled advantaged agents from the assimilation of high skilled agents. It follows that the optimal cutoff for assimilation is no longer consensual among agents with an advantaged background; rather, high-skilled agents favor less discrimination to induce more assimilation, and low-skilled agents favor more discrimination to induce less assimilation.

Suppose $d$ is chosen collectively by a finite subset $\mathcal{A}_F \subset \mathcal{A}$ of size $N$. Label these agents according to their ability, so that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_N$. Each $i \in \mathcal{A}_F$ strategically chooses $d_i \in [0, \bar{d}]$ for some arbitrary upper bound $\bar{d} \in \mathbb{R}_+$, and the vector $(d_1, \ldots, d_N)$ aggregates into a difficulty of assimilation $d \in \mathbb{R}_+$. We need not specify exactly how this aggregation takes place: $d$ can be he minimum of all the individual $d_i$ values, or the maximum, or the median, or any other order-statistic. Suppose that for some integer $n \in \{1, \ldots, N\}$, $d$ is the $n$th largest component of $(d_1, \ldots, d_N)$. The intuition is that at least $n$ agents must support erecting a given barrier to assimilation in order for this barrier to materialize. We then find that in equilibrium, the chosen discrimination level $d_i$ is strictly decreasing in $\theta_i$: highly able agents want their group to discriminate less; their less able co-ethnics want their group to discriminate more. If instead we assume $d = \frac{1}{N} \sum_{i=1}^{N} d_i$, agents with an advantaged background and skill below some level discriminate maximally $d_i = \bar{d}$ because they find their group is too open to assimilation; while those with skill above this level do not discriminate at all $d_i = 0$ because they find that their group already discriminates too much.

This prediction is testable. For instance, in the context of the United States, it predicts that more educated non-Hispanic white citizens hold more favorable attitudes toward ethnic minorities and toward immigrants than their less educated non-Hispanic white co-ethnics.

**Testable Hypothesis 7:** *Education is positively correlated with more favorable attitudes toward immigrants and ethnic minorities.*

The evidence on the relation between education and attitudes toward immigrants is consistent with this prediction. Using data from the 2003 European Social Survey, Hainmueller and Hiscox (2007) find that in Europe “people with higher levels of education and occupa-
tional skills are more likely to favor immigration” as predicted. The effect of education is still positive, though smaller, in poorer countries (see Figure 3 in the cross-national study by Mayda 2006).

f) Endogenous and collective choice of the cost of peer-pressure

Results are robust if we endogenize the cost of peer pressure $K$. For each $J \in \{A, D\}$, let $l_J$ choose not only punishment threshold $s^P_J$ but also punishment cost $K_J \in [K_-, K^+]$ with $K_- < 0 < K^+$, and assume that any $i \in J$ who chooses $s_i > s^P_J$ incurs punishment $K_J$. Under this extension, in equilibrium $l_D$ chooses $K_D = K^+$ (the proof is available from the author), so there is no loss in directly assuming that $K_D = K$ with $K > 0$ exogenous.

We can also generalize the theory to recognize that in practice, peer pressure is exercised by a group, not by a single individual. Let the choice of peer punishments be a collective one. For each $J \in \{A, D\}$, let $M_J$ be a finite set of agents with background $J$ and ability less than $\frac{1}{2}$, and assume each agent $i \in M_J$ chooses an individual punishment cutoff $s^P_i$, and that the group punishment $s^P_J$ is obtained as the aggregation of all these individual cutoffs ($s^P_J$ can be the average, or the median, or any other order statistic). Since all agents in $M_J$ share a common incentive to maximize the average skill $s_J$, for any equilibrium of the game in which a single pair of agents $l_D$ and $l_A$ respectively choose $(s^P_D)^* < 1$ and $(s^P_A)^* = 1$, it is immediate to construct an analogous equilibrium of the game with a collective choice of punishment thresholds in which $s^P_i = (s^P_D)^*$ for any $i \in M_D$ and $s^P_i = 1$ for any $i \in M_A$.

If, in a further extension, we assume that it is costly to exert pressure, then generating the desired peer-pressure becomes a collective-action problem for the agents with a disadvantaged background. Assume that any $i \in M_D$ choosing $s^P_i < 1$ bears a cost of exercising peer pressure if peer pressure takes place. This collective-problem can be solved in standard ways: if $k$ agents in $M_D$ are necessary to sustain peer pressure, and if the cost of punishing is not too high relative to the utility from skill, there is an equilibrium in which exactly $k$ agents in $M_D$ choose $s^P_i = (s^P_D)^*$. If the cost of punishing is too high to sustain collective action in this simple manner, it can be sustained as in Peski and Szentes (2013): an equilibrium with social sanctions against those who posses an arbitrary trait can be sustained if those who deviate and do not sanction become tainted by association and are themselves subsequently
sanctioned. This social sanctions are easier to sustain in our application, because sanctions are not arbitrarily applied against a payoff-irrelevant trait (as in Peski and Szentes 2013), but rather they are chosen strategically to align incentives in such a way that the sanctioning equilibrium is Pareto-improving for the agents who carry out the sanctions (see Example 1).

5 Policy Implications

I have theorized that in the absence of government intervention, the most able and most qualified members of poor ethnic minorities or immigrant groups assimilate, while their less talented peers are left behind in their own segregated niches of society. This outcome benefits the most advantaged group, by enriching it with highly skilled new members, but it splits disadvantaged communities into two subgroups: the successful few who assimilate, and the impoverished many who are left behind.

Affirmative action policies targeted at the most successful such as preferential college admissions sharpen this divide between winners and losers among those with a disadvantaged background: “enforcement of affirmative action guidelines was beneficial, but only to more qualified blacks” (Son, Model and Fisher 1989). More generally, “a variety of (...) shifts have brought great progress to middle-class blacks, (...) pushing a less qualified black population into an increasingly isolated and alienated underclass” (Son, Model and Fisher 1989 paraphrasing Wilson 1978).

If integration or assimilation across the socio-economic spectrum is deemed normatively desirable, as it was in the US or France in the first half of the twentieth-century (Brubaker 2001), affirmative action or other public interventions to foster integration must target and benefit the weakest, least skilled members of underprivileged communities; these agents, who are least welcome in mainstream society, are those who most need favorable policies to integrate. The Moving to Opportunity for Fair Housing program by the US Dept. of Housing and Urban Development is one policy intervention that facilitates assimilation by the least skilled; other possible policies would be tax incentives for companies in low-pay sectors to employ a diverse and integrated workforce. These policies better incentivize integration at
the bottom of the skill distribution than affirmative action policies targeted at the top of the skill distribution.

If assimilation is not deemed normatively desirable, and voluntary segregation in a multicultural society is regarded as satisfactory, a different normative consideration arises: in the absence of public policy intervention, the disadvantaged community is impoverished, and the advantaged community enriched, as a direct consequence of the transfer of talent across communities through assimilation. An argument to favor redistributive policies that tax the advantaged community and subsidize the disadvantaged one could be that these policies compensate the disadvantaged community for its loss of talent.

With regard to the acting white phenomenon, the policy recommendation to tackle it is clear: create incentives so that students become stakeholders in the success of their most able classmates.

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

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

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

6 Appendix

Lemma 1

Proof. For each $J \in \{A, D\}$, let $s_J(x)$ and $s_D(x)$ are the average skill in $A$ and the average skill in $D$ as a function of $x$ assuming that agents in $D$ assimilate if and only if their skill is above $x$. Then

$$s_A(x) = \left(\frac{1}{2} + \frac{g + x - g}{2 \frac{g}{x}}\right) \frac{1}{1 + \frac{g-x}{g}} = \frac{g + (g + x)(g - x)}{2(2g - x)}.$$  

The first order condition for maximization of $s_A(x)$ is

$$\frac{1}{2 (x - 2g)^2} (g^2 - 4gx + g + x^2) = 0$$

$$g^2 - 4gx + g + x^2 = 0.$$  

The only solution in the interval $[\frac{1}{2}, 1]$ for this condition is $x = 2g - \sqrt{g(3g - 1)}$. Because the first derivative of $s_A(x)$ is continuous, positive for $x = \frac{1}{2}$ and negative for $x = 1$ and zero only for $x = 2g - \sqrt{g(3g - 1)}$ within interval $[\frac{1}{2}, 1]$, it follows that the second order condition is satisfied as well and

$$x^*(g) = 2g - \sqrt{g(3g - 1)}.$$  

The average skill level $s_A = 2g - \sqrt{g(3g - 1)}$ is achieved if and only if (up to any deviation by a mass zero of agents, who have no effect on averages) every $i \in A$ chooses maximal skills according to her potential ($s_i = \theta_i$), every $j \in D$ with $\theta_i > 2 - \sqrt{3 - 1/g}$ chooses skill $s_j = g\theta_j$ (so that $s_j > 2g - \sqrt{g(3g - 1)}$) and assimilates, and no $j \in D$ with $\theta_i < 2 - \sqrt{3 - 1/g}$
assimilates. □

**Proposition 2**

**Proof.** First step. Show that in equilibrium, every $i \in A$ chooses $s_i = \theta_i$ and every $i \in D$ chooses $g\theta_i$.

Note that since $K = 0$, the choice of $s^P_J$ is not payoff-relevant. Assume any values are chosen. At the second stage, each agent $i$ chooses $s_i$. Since $s_i$ is private information it does not affect future play by any other agent, and since the utility for $i$ is ceteris paribus strictly increasing in $s_i$, it follows that it is strictly dominated for any agent $i$ to choose any $s_i$ other than the maximum $i$ can attain. Hence every $i \in A$ chooses $s_i = \theta_i$ and every $i \in D$ chooses $g\theta_i$.

Second step. In equilibrium, there is a cutoff for assimilation.

At the third stage, agents in $D$ 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 the upper bound of her interval of skill choices.

Given any $d$ and any assimilation profile $a_{-i}$ for every $j \in D \setminus \{i\}$, agent $i$ chooses $a_i = 1$ if and only if $s_i$ is above some cutoff that depends on $d$ and $a_{-i}$. For any $i,j \in D$ such that $s_i > s_j$, and given any $d$ and any strategy profile $a_{-i,j}$ for every $h \in D \setminus \{i,j\}$, if $i$ and $j$ best respond, $a_j = 1$ implies $a_i = 1$. Hence for any $d \in \mathbb{R}_+$, there is a cutoff $\tilde{\sigma} \in [0,1]$, which depends on $d$, such that any $i \in D$ strictly prefers to choose $a_i = 1$ if and only if $s_i$ is strictly above $\tilde{\sigma}$, (weakly above for the special corner case in which every agent strictly prefers to assimilate and $\tilde{\sigma} = 0$). I show below that the solution for the value of the cutoff $\tilde{\sigma}$ is in equilibrium unique, but until I prove this I let $\sigma : \mathbb{R}_+ \rightarrow [0,1]$ denote the set of all possible cutoffs for assimilation that constitute an equilibrium of the continuation game after $d$ is chosen and observed by disadvantaged agents, so that $\sigma(d)$ is the set of possible cutoffs for a given $d$.

Third step. Identify the cutoff that generates the best outcome for $A$.

For any $x \in [0,1]$, there exists a unique value of $d$ such that $i \in D$ with $s_i = x$ is indifferent between assimilating or not given that other agents assimilate if and only if their
skill is above $x$. Let it be denoted $\sigma^{-1}(x)$. The optimal cutoff of assimilation for agent $h \in \mathcal{A}$ is $x^* = \arg \max_{x \in [0,g]} s_h s_A(x)$ s.t. $s_A(x) = \frac{g + (g-x)(x+g)}{2(2g-x)}$, which is $\arg \max_{x \in [0,g]} \frac{g + g^2 - x^2}{2(2g-x)}$. This optimization problem is the maximization of a continuous function over a compact set, hence it has a solution $x^*$. Then there exists an equilibrium in which $h \in \mathcal{A}$ chooses $d = d^* = \sigma^{-1}(x^*)$, every $i \in \mathcal{D} \cup \mathcal{A}$ chooses $s_i = \theta_i$ and every $i \in \mathcal{D}$ chooses $a_i = 1$ if and only if $s_i > x^*$, or, equivalently, if and only if $\theta_i > \theta^* = \frac{1}{g} x^*$. It follows directly from Lemma 1 that

$$x^*(g) = 2g - \sqrt{g(3g-1)} \quad \text{and} \quad \theta^*(g) = 2 - \sqrt{3 - 1/g}$$

are the solutions for skill and ability cutoffs, expressed as functions of the exogenous parameter. It follows that $\theta^* \in \left(\frac{1}{2}, 1\right)$ for any $g \in \left(\frac{1}{2}, 1\right)$.

Fourth step. Uniqueness.

To prove uniqueness, we calculate the equilibrium level of discrimination $d^*$.

Note that

$$s_A(x^*(g)) = s_A^*(g) = \frac{g + (g - (2g - \sqrt{g(3g-1)}))(g + (2g - \sqrt{g(3g-1)}))}{2(2g - (2g - \sqrt{g(3g-1)}))} = \frac{g + 2g\sqrt{g(3g-1)} - 3g^2}{\sqrt{g(3g-1)}} = 2g - \sqrt{g(3g-1)}, \quad \text{and}$$

$$s_D(x^*(g)) = s_D^*(g) = g - \frac{1}{2} \sqrt{g(3g-1)},$$

so in order for $i$ with $s_i = x^*$ to be indifferent about assimilation, it must be

$$x^*(s_A(s^*) - s_D(s^*)) = d(1 - x^*), \quad \text{or}$$

$$\left(2g - \sqrt{g(3g-1)}\right) \left(2g - \sqrt{g(3g-1)} - \left(g - \frac{1}{2} \sqrt{g(3g-1)}\right)\right) = d \left(1 - \left(2g - \sqrt{g(3g-1)}\right)\right),$$

so we can express the equilibrium discrimination as a function of $g$ thus:

$$d^*(g) = \frac{2g - \sqrt{g(3g-1)}}{1 - \left(2g - \sqrt{g(3g-1)}\right)} \left(g - \frac{1}{2} \sqrt{g(3g-1)}\right)^2 = \frac{2 \left( g - \frac{1}{2} \sqrt{g(3g-1)}\right)^2}{1 - \left(2g - \sqrt{g(3g-1)}\right)}.$$
Next define \( b(x) = x(s_A(x) - s_D(x)) \) as the benefit of assimilation function, which measures the benefit of assimilating for the agent at the cutoff \( x \), as a function of the cutoff, if agents in \( \mathcal{D} \) assimilate if and only if their type is above the cutoff \( x \). The function \( b(x) \) is

\[
x(s_A(x) - s_D(x)) = x \left( \frac{g + (g - x)(g + x)}{2(2g - x)} - \frac{x}{2} \right).
\]

Its first derivative is

\[
\frac{\partial}{\partial x} \left( x \left( \frac{g + (g - x)(g + x)}{2(2g - x)} - \frac{x}{2} \right) \right) = \frac{g}{(x - 2g)^2} \left( g^2 - 4gx + g + x^2 \right).
\]

Its second derivative is

\[
\frac{\partial}{\partial x} \left( \frac{g}{(x - 2g)^2} \left( g^2 - 4gx + g + x^2 \right) \right) = 2g^2 \frac{3g - 1}{(x - 2g)^3} < 0
\]

so the function is concave.

Given that the benefit function \( b(x) \) is concave, and the cost of assimilation function \( c(d, s_i) = d(1 - s_i) \) evaluated at \( s_i = x \) is linearly decreasing in \( x \), it follows that for any \( d \), the net benefit of assimilating \( b(x) - c(d, x) \) is either always increasing in \( x \) (in which case it can be zero only once), or first increasing and then decreasing (in which case it could be zero once, or twice at two different equilibria). Suppose \( b(x) - c(d, x) > 0 \) at \( x = g \) and \( d = d^*(g) \); then even if \( b(g) - c(d^*(g), g) \) is first increasing and then decreasing, it can only be zero once, and the continuation game that follows after advantaged agent \( h \) chooses \( d^*(g) \) has a unique equilibrium at this value of \( x \) such that \( b(x) = c(d^*(g), x) \). Since this is the unique first best for agent \( h \), it follows that \( h \) indeed chooses \( d \). Thus we want to show that \( b(x) - c(d, x) > 0 \) at \( d = d^*(g) \) and \( x = g \).

Note that

\[
b(g) - c(d^*(g), g) = g \left( \frac{g + (g - g)(g + g)}{2(2g - g)} - \frac{g}{2} \right) - \frac{2 \left( g - \frac{1}{2} \sqrt{g(3g - 1)} \right)^2}{1 - \left( 2g - \sqrt{g(3g - 1)} \right)} (1 - g)
\]

34
\[= g \frac{(1-g)}{2} - \frac{2 \left( g - \frac{1}{2} \sqrt{g(3g-1)} \right)^2}{1 - \left( 2g - \sqrt{g(3g-1)} \right)} (1-g). \]

So we want to show
\[
\frac{g}{2} - \frac{2 \left( g - \frac{1}{2} \sqrt{g(3g-1)} \right)^2}{1 - \left( 2g - \sqrt{g(3g-1)} \right)} > 0
\]
\[
g \left( 1 - 2g + \sqrt{g(3g-1)} \right) > \left( 2g - \sqrt{g(3g-1)} \right)^2
\]
\[
1 - 2g + \sqrt{g(3g-1)} > \left( 2\sqrt{g} - \sqrt{3g-1} \right)^2
\]
\[
1 - \sqrt{g} \left( 2\sqrt{g} - \sqrt{3g-1} \right) > \left( 2\sqrt{g} - \sqrt{3g-1} \right)^2
\]
\[
1 + \left( 2\sqrt{g} - \sqrt{3g-1} - \sqrt{g} \right) \left( 2\sqrt{g} - \sqrt{3g-1} \right) > 0
\]
\[
1 + 2g - 3\sqrt{g} \sqrt{3g-1} + 3g - 1 > 0
\]
\[
5g - 3\sqrt{g} \sqrt{3g-1} > 0
\]
\[
5\sqrt{g} > 3\sqrt{3g-1},
\]
or equivalently \(3 > 2g\), which holds for any \(g \in \left( \frac{1}{2}, 1 \right)\), so \(b(x) - c(d, x) > 0\) at \(x = g\) and \(d = d^*(g)\) for any \(g \in \left( \frac{1}{2}, 1 \right)\) as desired, and thus there is only one equilibrium of the continuation after \(d^*\) is chosen, and thus \(d^*\) is indeed chosen in the unique equilibrium of the whole game. Therefore, the equilibrium outcome is unique. \(\blacksquare\)

**Proposition 3**

**Proof.** I first prove the existence claim.

Let \(\Omega = (\Omega_D, \Omega_A)\) be a pair of distributive functions of the skill in \(D\) and \(A\). Wlog., constrain \(s^P_D\) to be chosen in \([0, g]\).

Note first that at the second stage, for any \(i \in J \in \{A, D\}\), since \(s_i\) is private information, it has no effect on the actions taken by any other player at stage 3, and since \(s_is_J - a_id(1-s_i)\) is strictly increasing in \(s_i\), it follows that ceteris paribus the utility of \(i\) is strictly increasing in \(s_i\) for \(s_i \in [0, s^P_J]\) and for \(s_i \in (s^P_J, 1]\) with a discontinuity at \(s^P_J\). It follows that choosing any \(s_i \notin \{\theta_i, s^P_J\}\) is strictly dominated either by \(s_i = \theta_i\) or by \(s_i = s^P_J\). Therefore, in equilibrium \(s_i \in \{s_i, s^P_J\}\), hence \(\Omega_J\) has uniform density on \([0, s^P_J]\) and positive mass only at \(s^P_J\).
At the third stage, by an analogous argument as in the proof of Proposition 2, there is a cutoff \(s(d, \Omega) \in [0, 1]\) such that agent \(i \in \mathcal{D}\) chooses \(a_i = 1\) if \(s_i > s(d, \Omega)\) and chooses \(a_i = 0\) if \(s_i < s(d, \Omega)\). Unlike in the proof of Proposition 2, the cutoff may not be unique; if it is not unique, arbitrarily select the solution with the highest cutoff and fewest agents assimilating.

At the second stage, in anticipation of the equilibrium in stage 3, any agent \(i \in \mathcal{A}\) with \(\theta_i \leq s^P_A\) uniquely best respond by choosing \(s_i = \theta_i\). Any agent \(i \in \mathcal{A}\) with \(\theta_i > s^P_A\) faces a trade-off: choosing \(s_i = \theta_i > s^P_A\) she incurs a cost \(K\), but she derives a benefit \((\theta_i - s^P_A)s_A\).

The benefit of choosing \(s_i = \theta_i\) is increasing in \(\theta_i\), while the cost is fixed at \(K\). Thus, there is a cutoff \(\theta(s^P_A, K)\) such that \(\theta(s^P_A, K) > s^P_A\) and such that any \(i \in \mathcal{A}\) with \(\theta_i > \theta(s^P_A, K)\) chooses \(s_i = \theta_i\) and any \(i \in \mathcal{A}\) with \(\theta_i \in (s^P_A, \theta(s^P_A, K))\) chooses \(s_i = s^P_A\).

Similarly, any agent \(i \in \mathcal{D}\) with \(\theta_i \leq \frac{1}{g} s^P_D\) uniquely best respond by choosing \(s_i = g\theta_i\). Any agent \(i \in \mathcal{D}\) with \(\theta_i > \frac{1}{g} s^P_D\) faces a trade-off: choosing \(s_i = g\theta_i > s^P_D\) she incurs a cost \(K\), but she derives a benefit \((\theta_i - s^P_D)s_D\), and a reduced cost of assimilation if she assimilates. The benefit of choosing \(s_i = g\theta_i\) is increasing in \(\theta_i\), while the cost is fixed at \(K\). Thus, there is a cutoff \(\theta(s^P_D, K, d)\) such that \(\theta(s^P_D, K, d) > \frac{1}{g} s^P_D\) and such that any \(i \in \mathcal{D}\) with \(\theta_i > \theta(s^P_D, K, d)\) chooses \(s_i = g\theta_i\) and any \(i \in \mathcal{A}\) with \(\theta_i \in \left(\frac{1}{g} s^P_D, \theta(s^P_D, K, d)\right)\) chooses \(s_i = s^P_D\).

At the first stage, note first that in equilibrium \((s^P_A)^* = 1\). Choosing \(s^P_A < 1\) causes any \(i \in \mathcal{A}\) with \(\theta_i \in (s^P_A, \theta(s^P_A, K))\) to choose skill \(s_i = s^P_A < \theta_i\), which reduces \(s_A\), making \(l_A\) strictly worse off.

For any \(s^P_D \in [0, g]\) and any \(d \in \mathbb{R}_+\) and given \(K\), let \(f(s^P_D, d, K)\) denote the ability of the marginal agent \(i \in \mathcal{D}\) who is indifferent between \(s_i = g\theta_i\) and making the optimal assimilation decision, or \(s_i = s^P_D\) and making the optimal assimilation decision, given that any \(i \in \mathcal{D}\) with \(\theta_i \leq s^P_D\) chooses \(s_i = g\theta_i\) and \(a_i = 0\), every \(i \in \mathcal{D}\) with \(\theta_i \in \left[\frac{1}{g} s^P_D, f(s^P_D, d, K)\right)\) chooses \(s_i = s^P_D\) and \(a_i = 0\), and every \(i \in \mathcal{D}\) with \(\theta_i \geq f(s^P_D, d, K)\) chooses \(s_i = g\theta_i\) and \(a_i = 1\). If no agent is indifferent, let \(f(s^P_D, d, K) = 1\).

Assume that indeed every \(i \in \mathcal{D}\) with \(\theta_i \leq \frac{1}{g} s^P_D\) chooses \(s_i = g\theta_i\) and \(a_i = 0\), every \(i \in \mathcal{D}\) with \(\theta_i \in \left[\frac{1}{g} s^P_D, f(s^P_D, d, K)\right)\) chooses \(s_i = s^P_D\) and \(a_i = 0\), and every \(i \in \mathcal{D}\) with \(\theta_i \geq f_2(d)\) chooses \(s_i = g\theta_i\) and \(a_i = 1\). Then, given these assumptions, we can consider \(s_A : [0, 1] \rightarrow [0, 1]\) as a function of \(s^P_D\) and then notice that \(s_A(s^P_D) - s^P_D\) is a continuous
function that is strictly positive for \( s_D^P \) sufficiently low and strictly negative for \( s_D^P \) sufficiently high, thus, by the intermediate value theorem, there exists a value such that \( s_A(s_D^P) - s_D^P = 0 \). I construct an equilibrium in which \( (s_D^P)^* = s_A(s_D^P) \) and \( d^* \) is such that every agent \( i \in D \) with \( s_i = (s_D^P)^* \) is indifferent between assimilation or not, and the assumptions at the beginning of this paragraph indeed hold.

I now check that there is no profitable deviation. By construction, at stage 3, \( i \in D \) with \( s_i = (s_D^P)^* \) is indifferent about assimilation, so in equilibrium does not assimilate, which also implies that any \( i \in D \) with \( s_i > (s_D^P)^* \) chooses \( a_i = 1 \) and any \( i \in D \) with \( s_i < (s_D^P)^* \) chooses \( a_i = 0 \) as assumed. At stage 2, every \( i \in A \) chooses \( s_i = \theta_i \), which is optimal given that at stage 1 \( (s_A^P)^* = 1 \). Every agent \( i \in D \) with \( \theta_i \leq \frac{1}{q} s_D^P \) chooses \( s_i = g \theta_i \) as assumed, which is optimal; every \( i \in D \) with \( \theta_i \in \left[ \frac{1}{g} s_D^P, f((s_D^P)^*, d^*, K) \right] \) chooses \( s_i = s_D^P \), which is again optimal by definition of \( f(s_D^P, d, K) \); and every \( i \in D \) with \( \theta_i \geq f((s_D^P)^*, d^*, K) \) chooses \( s_i = g \theta_i \), which is optimal as implied by the definition of \( f(s_D^P, d, K) \), so the assumptions at the second stage are verified. At stage 1, we have already argued that \( (s_A^P)^* = 1 \). We now note that choosing \( s_D^P < (s_D^P)^* \) is not a profitable deviation: it makes at least some agents with a disadvantaged background and \( \theta_i \in (\frac{1}{g} s_D^P, \frac{1}{g} (s_D^P)^*) \) to choose \( s_i = s_D^P < g \theta_i \), which reduces the average skill level \( s_D^P \); choosing \( s_D^P > (s_D^P)^* \) is not a profitable deviation because it induces all agents with a disadvantaged background and \( \theta_i \in (\frac{1}{g} (s_D^P)^*, \frac{1}{g} s_D^P) \) to choose \( s_i = g \theta_i \) and \( a_i = 1 \), which reduces \( s_D^P \); choosing \( d < d^* \) causes all \( i \in D \) with \( \theta_i = s_D^P \) and some \( i \in D \) with \( \theta_i < s_D^P \) to choose \( a_i = 1 \), which reduces the average skill \( s_A \); and choosing \( d > d^* \) is again not profitable, as it either has no effect, or it causes all agents \( i \in D \) with \( \theta_i \in f((s_D^P)^*, d^*, K), f((s_D^P)^*, d, K) \) to choose \( a_i = 0 \), which either has no effect if \( f((s_D^P)^*, d^*, K), f((s_D^P)^*, d, K) \) is empty (because \( f((s_D^P)^*, d^*, K) = f((s_D^P)^*, d, K) \)), or else it reduces the average skill \( s_A \). Thus, there is no profitable deviation at any stage, and the strategy profile constitutes an equilibrium.

Next I prove the claim that in every equilibrium \( s_D^P < g \) and \( s_A^P = 1 \).

First note that as mentioned above, choosing \( s_A^P < 1 \) causes any \( i \in A \) with \( \theta_i \in (s_A^P, \theta(s_A^P, K)) \) to choose skill \( s_i = s_A^P < \theta_i \), which reduces the average skill level \( s_A \), making \( l_A \) strictly worse off.
Suppose $s_P = g$. Then, the unique equilibrium with $s_P = g$ and $s_A = 1$, is such that any $i \in D$ with $\theta_i > \theta^* = 2 - \sqrt{3 - 1/g}$ assimilates. Suppose $l_D \in D$ (the agent who chooses $s_P$) deviates to choose $s_P = g\theta^*$. Then any $i \in D$ with $\theta_i \in (\theta^*, f((s_P), d^*, K))$ would choose $\theta_i = s_P$ and $a_i = 0$ which would then increase the average skill level $s_D$, thus benefit $m$, so the original profile with $s_P = g$ is not an equilibrium. Thus, in every equilibrium $s_P < g$. ■

References


