Stereotypes and Identity Choice

Young-Chul Kim[*†] Glenn C. Loury[‡]

January 22, 2015

Abstract

We extend an economic model of 'collective reputations' (sometimes referred to as 'stereotyping') by allowing observed agents to exert control over their perceived identities. The logic of individuals' identity choices exacerbates group disparities by inducing a positive selection of the more talented individuals into a group with a superior reputation. Thus, the inequality deriving from the stereotyping of endogenously constructed social groups is at least as great as the inequality that can emerge when perceived identity is not malleable. Among the human behaviors potentially illuminated by this theory are: (1) the selective out-migration from a stigmatized group associated with 'passing' and (2) the production of the indices of differentiation by better-off members of the negatively stereotyped group. We prove that these identity manipulation activities can increase the total welfare of the society under some limited conditions, though they may generate a conflict among the stereotyped population because of their adverse impact on the reputations of those who are left behind.

Keywords: Stereotypes, Statistical Discrimination, Identity Choice, Inequality.
JEL Codes: D63, J15, J70, Z13

[*]We thank the seminar/workshop participants for valuable comments and discussions at MIT, the Santa Fe Institute, CEP/LSE Labour Seminar (London School of Economics and Political Science), the Becker Friedman Institute (Univ of Chicago), Brown Univ, Univ of East Anglia, Federal Reserve Bank of St. Louis, Sogang Univ and so on. We are also grateful to the conference participants for helpful comments and suggestions, which include the 4th World Bank Group Conference on Equity (2014), the PET Workshop on the Political Economy of Development (2011), the Asian Meeting of Econometric Society (2011) and the Annual Meeting of Korean Association of Public Finance (2014). We are indebted to Rajiv Sethi, Matthew Jackson, Alan Kirman, Samuel Bowles, Yeon-Koo Che and Biung-Ghi Ju for their thoughtful and insightful comments. All remaining errors are ours.

[†]Department of Economics and Finance, Sangmyung University, Jongno-gu, Seoul 110-743, South Korea. Email: yckim.kdi@gmail.com.

[‡]Department of Economics, Box B, Brown University, Providence, RI 02912, USA. Email: Glenn_Loury@brown.edu.

1 Introduction

Social information is valuable, and many people seek it in daily life. One of the ways that we generate and store social information is to classify the persons we encounter, i.e., form broad categories between which contrasts can be drawn and about which generalizations can be made. Through classification, we can better understand what is to be expected from those with whom we must interact but about whom all too little can be discerned. The information-hungry observers, in making pragmatic judgments, have such an incentive to use group-average information to assess a subject's functionally relevant traits when they are not directly observable. Thus, the practice of grouping people together on the basis of their common possession of visible marks or other observable characteristics is a universal aspect of the human condition. The 'collective reputations' are this sort of rational formation by external observers of beliefs about the unobserved traits of varied population aggregates. The phenomenon, sometimes referred to as 'stereotyping', has long been of interest to economists (e.g., Arrow, 1971; Coate and Loury, 1993; Tirole, 1994; Fang, 2001), sociologists (e.g., Goffman, 1959; Anderson, 1990; Sampson and Raudenbush, 2004), and social psychologists (e.g., Fiske, 1998; Greenwald and Banaji, 1995; Steele and Aronson, 1995). In this paper, we extend the economics literature about collective reputations and stereotypes by allowing observed agents to exert control over their perceived identities.

When a stranger comes into our presence, first appearances are likely to enable us to anticipate his category and attributes, though the true attributes he could, in fact, possess are different from the anticipated ones (Goffman, 1963). This implies a fundamental distinction between social identity, which addresses how an individual is perceived and categorized by others, and personal identity, which is the distinct personality of an individual regarded as a persisting entity (Tajfel, 1974). An individual's success in everyday life can be influenced substantially by the social identity attached to him. Then, incurring some cost, individuals may take actions that affect the way in which they are categorized and perceived by observers. For instance, one may decide whether to smoke or not or whether to dress formally or casually if he believes that this specific categorization influences the way he is treated by others. Therefore, the choice of perceived social identity is a rational behavior of economic agents in societal settings.

Developing an identity choice model, we use a stereotyping-cum-signaling framework pio-

neered by Arrow (1973) and Coate and Loury (1993). When a job candidate's productivity is not perfectly observable, employers in the screening process have an incentive to use the collective reputations of the identity groups to which the job applicants belong. This can generate multiple self-confirming prior beliefs on the part of employers about different social identity groups. Individuals belonging to a group with a better collective reputation have a greater incentive to acquire the attributes valued in the marketplace than do those who belong to a group with a poor reputation. However, given its greater acquisition rate of valued attributes, the group can maintain this better collective reputation. On the other hand, individuals belonging to a group with a poor collective reputation have a smaller incentive to acquire the valued attributes, and with the lower acquisition rate, the employers' negative stereotype against this group is also self-confirmed. Therefore, in this framework, the multiple self-confirming beliefs explain the inequality of collective reputations between exogenous and equally endowed identity groups as being due to the positive feedback between a group's reputation and its members' investment incentives.

We extend this set of arguments by relaxing the immutability assumption: Instead of exogenously given identities, people are able to control how they are categorized or perceived by others. If they are different in terms of economically relevant dimensions such as ability or productivity and if they anticipate that one type of identity will be better treated than another in the marketplace, the incentive for people to join the favored group varies according to the ability. The identity choice behaviors will systematically induce a positive selection along the ability parameter in the group that is anticipated to be better treated. The result is that human capital cost distributions between groups endogenously diverge, which reinforces incentive-feedbacks. This creates an additional type of self-fulfilling prophecy that can generate inequality between identity groups, which is a different mechanism than the positive complementarities between collective reputation and skill investment incentives. When these two mechanisms, positive selection and positive complementarities, are jointly operative, we have greater inequality between two identity groups than would have been the case in the absence of the endogeneity of identity choice.

There are many situations in which identity choice and group stereotypes operate in tandem. Among the behaviors potentially illuminated by our theory are: (1) the selective 'out-migration' from a stigmatized group associated with 'passing' and (2) the production of the indices of differentiation by better-off members of the negatively stereotyped group (e.g., affec-

tations of speech, dressing up and conspicuous consumption). In the following, we introduce the possible rationale behind those behaviors and depict the various circumstances that are relevant for each phenomenon.

Stigmatized groups are identified in various ways around the world: along racial lines in societies such as the United States, South Africa, Australia and many Latin American countries, along religious lines in Pakistan, Northern Ireland and Israel, along ethnic lines in Singapore, Indonesia and the Balkan states, with caste-like social division in the Indian sub-continent and the treatment of Gypsies and immigrants in Europe. Young members in a stigmatized group may consider "passing" into the better-regarded group when the return for "passing" (e.g., better treatment in the labor market) outweighs its cost (e.g., loss of ties to one's own kind.) The selective out-migration occurs as more talented members in the disadvantaged groups cross the color/religious/ethnic/caste lines disproportionately.

A well-known example is the ethnic Koreans in Japan (referred to as "Zainichi"), many of whom are descended from forced laborers in mines and factories who were brought to Japan from the Korean peninsula during the period of Japanese imperialism. Every year, approximately 10,000 Koreans, of approximately 600,000 Korean descendants holding Korean nationality, choose to be naturalized as 'official' Japanese, giving up their names and original nationality. To escape the negative stereotypes and prejudices against the Zainichi, many of the naturalized Koreans conceal their ethnicity, pretending that they have no knowledge about Korean culture and language (Fukuoka et al., 1998). As disproportionately more talented young Koreans pass for native Japanese when seeking formal employment or marriage, the inequality between the two identity groups (Zainichi and native Japanese) becomes persistent or even exacerbated.

Other than the Zainichi, who share a similar appearance with the Japanese, passing is harder for blacks in the United States due to their physical makeup. However, the light-skinned African-Americans with mixed ancestry have been crossing the boundaries of color and racial identity. According to the NLS79 National Longitudinal Survey conducted by the Department of Labor in the US, 1.87 percent of those who had originally answered "Black" in 1979 (when they were 14 to 22 years old) switched to answering the interviewer's race question with either "white," "I don't know," or "other" by 1998 (Sweet, 2004). In old Hollywood, for instance, talented movie stars were expected to downplay their ethnic origins when they were not solely of European extraction. Some of them successfully estranged themselves from

their roots and achieved fame and fortune in the moves, including Carol Channing (a quarter Black) and Merle Oberon (Anglo-Indian).

Unlike the United States, which had defined concepts of race due to the 'one drop rule,' whereby people with any known African ancestry were automatically classified as Black, racial classifications in Latin American and Caribbean countries are based primarily on skin tone and on other physical characteristics such as facial features, hair texture, etc. In these white supremacist societies, a dark skinned person is more likely to be discriminated against, and a light skinned person is considered more privileged (Telles, 2004). In their everyday life, the black-looking mixed race people tend to refuse to identify as Black, but the white-looking mixed race people gladly identify as White. Furthermore, the fascination with becoming "white" has increased over time with the prevailing practice of skin-bleaching and hair-straightening among the mixed-race youngsters.

In other situations, discriminated groups may modify their accents, word choices, manner of dress and even names in an attempt to appear to be members of a privileged group. For example, My Fair Lady, a musical based upon George Bernard Shaw's *Pygmalion*, concerns Eliza Doolittle, a Cockney flower girl who takes speech lessons from a phonetician so that she may pass as a lady in the high society of Edwardian London. This type of passing in the context of caste is called *Sanskritization*, which is a process by which a low or middle Hindu caste seeks upward mobility by emulating the rituals and practices of the upper or dominant castes. A caste may rise to a higher position in the hierarchy, in a generation or two, by adopting the Sanskritic theological ideas and the Brahminic way of life such as vegetarianism and teetotalism (Srinivas, 1952).

Passing into the better-regarded group is not always possible for every stigmatized group. For instance, it would be very hard when the pertinent physical traits are passed on across generations, are easily discerned and are not readily disguised. To inhibit being stereotyped, the most talented of the visibly distinct stigmatized population, who gain most by separating themselves from the mass, may develop the indices of differentiation that can send signals that they are different from the average of the stigmatized mass. Taking the example of the blacks in the United States, the strategies of social identity manipulation that can be adopted by better-off members are: affectations of speech, dressing formally rather than wearing casual clothes, spending more on conspicuous consumption and migration to affluent residential areas (Goffman, 1959). In short, these self-presentation methods for "partial passing" aim

to communicate "I'm not one of THEM; I'm one of YOU!" (Loury, 2002). Even when there is no priori difference in observable traits among the stereotyped population, we may see an autonomously growing sub-population with differentiated traits whose members are preferentially treated in the marketplace or in the law enforcement.

There is systematic empirical evidence regarding the styles of self-presentation for social identity manipulation. For instance, through the careful examination of the Consumer Expenditure Survey (CES) by the U.S. Department of Labor, Charles et al. (2009) report that blacks and Hispanics spend 30 percent more than similar whites on visible goods such as clothing, cars and jewelry. They conclude that blacks and Hispanics earning a higher income, who live in an area where the community income is relatively lower, have greater incentives to differentiate themselves and signal their high status by acquiring visible goods. Using audio data from interviews administered to the National Longitudinal Survey of Youth (NLSY) respondents, Grogger (2011) examines the role that speech pattern may play in explaining racial wage differences. He finds that, among blacks, speech patterns are highly correlated with the wages of young workers: Black speakers whose voices were distinctly identified as black earn approximately 12 percent less than whites with similar observable skills. This implies the more talented blacks tend to speak standard American English rather than African American English. Speech patterns may raise productivity. However, even when it does not, it may signal the worker's underlying skill. Then, the more talented in a stigmatized group will have greater incentives to learn the majority language.

The identity choice model in this paper explores implications of the fact that the distribution of abilities within distinct identity groups becomes endogenous when individuals choose how they will be identified by external observers. We start with a standard statistical discrimination framework based on Coate and Loury's (1993) original setup, which helps us identify the multiple self-confirming prior beliefs of employers about the different social groups with identical fundamentals. This standard framework entails no selection into or out of the groups, implying that human capital cost distributions are identical across groups in equilibrium. We call the self-confirming belief equilibrium with exogenous social identities Phenotypic Stereotyping Equilibrium (PSE), using the term 'phenotype' to indicate exogenously determined immutable appearance. When membership is endogenous, however, the better-regarded group will, in equilibrium, come to consist disproportionately of high ability/low human capital investment cost types because they are the ones who gain most from

joining a favored group. We call such a group-disparate equilibrium with endogenous identities an Endogenous Stereotyping Equilibrium (ESE).

For the development of our theoretical model, we introduce two postures (affects A and B) indicating that people can make a choice regarding how they will be categorized and perceived in the marketplace. Every single individual not only decides whether to acquire skills but also elects to present himself as affect A or as affect B. Agents are endowed with two cost parameters, c and k, which represent the human capital investment cost and the identity manipulation cost (the cost of being perceived as affect A rather than B). We assume that the two parameters are statistically independent of each other, implying that a person's identity orientation cannot be used to predict their economic abilities. Nevertheless, we demonstrate that low human capital cost types are disproportionately drawn to the favored group with a superior collective reputation. In this regard, we are inspired by Fang's (2001) examination of the economic meaning of social culture, in which he indicates that a skilled worker can be more willing than an unskilled worker to undertake a specific cultural activity in a cultural equilibrium.

Comparing PSE and ESE in the given identity choice model, we find that, while inequality in PSE is due to the positive feedback between the reputation and investment incentive, inequality in ESE is due to the positive selection into the favored group as well as the reputation-incentive feedback. This ensures that the group inequality that derives from the environment in which people have options to migrate between categorized memberships is at least as great as the group inequality that can emerge from the phenotypic stereotyping in which the categorized membership is exogenously given and verifiable by everybody. We also prove the existence of such unequal ESE with endogenous identities, given the presence of multiple PSE with exogenous identities. In addition, those unequal ESE are the only stable equilibria when the society consists of a sufficiently large fraction of individuals whose identity manipulation cost is low, that is, the equality between groups is not sustainable when identity manipulation is easier to undertake.

Applying this theory to the passing and 'partial passing' phenomena, we find that non-passers (or non-partial passers) who are left behind are adversely affected by the selective out-migrations (or the usage of the indices of differentiations). The activities may undermine solidarity in a stereotyped population, as the worse-off members of the population accuse the passers (or partial passers) of some kind of immoral betrayal. This reasoning provides a

different explanation of the 'acting white' phenomenon from those offered by other scholars (e.g., Austen-Smith and Fryer, 2005).

However, through the decomposition of the societal efficiency gain into reputational externalities and passing (partial passing) premium, we prove that these identity manipulation activities can increase the total welfare of the society under some limited conditions, even when they generate a conflict among the stereotyped population. Furthermore, we demonstrate that when a stereotyped group is severely discriminated in the labor market equilibrium, the selective out-migration or the production of the indices of differentiation can improve the societal efficiency even without hurting the welfare of the left-behind.

The remainder of this paper is structured as follows. Section 2 describes the basic structure of the signaling model, in which agents decide on the perceived identity as well as the skill acquisition. Section 3 defines both the Phenotypic and Endogenous Stereotyping Equilibria. Section 4 studies the properties of the identity choice behaviors, and Section 5 examines the existence and the stability of Endogenous Stereotyping Equilibria. Section 6 follows with a discussion of the welfare properties of the equilibria, applying the theory to passing and 'partial passing' phenomena. Section 7 presents the study's conclusion.

2 Framework of the Model

In this section, we propose a collective reputation model in which group identities are endogenous. Imagine a large number of identical employers and a large population of workers, in which each employer is randomly matched to many workers. The workers not only make an investment decision on skill acquisition but also choose how to be perceived by others before they enter the labor market. The employers set the wage for each worker to be proportional to the worker's expected productivity using the productivity-related noisy signal as well as the perceived group identity. The following explains this in detail.[1]

First, we assume that the newborn agents choose whether to be skilled before they enter the labor market: $e \in \{0, 1\}$. The cost of obtaining a skill varies among the newborns: $c \in [0, \infty]$. Agents with less cost are more capable individuals, and they can acquire skills more easily. Let $G(c)$ be the fraction of workers with a skill acquisition cost no greater than

[1]The basic structure of this signaling framework is introduced first by our earlier work, Kim and Loury (2012). This paper is the complete development of the long-term research project on stereotypes and identity choice.

c. The cost distribution $G(c)$ satisfies $G(0) > 0$ and $G(\infty) = 1$, which imply the existence of a fraction of highly capable (incapable) workers whose skill acquisition cost is sufficiently low (high). We impose that the related density function of the cost, $g(c)$, is a single-peaked function of c, increasing (decreasing) for any c less (greater) than $\hat{c}$. Workers must decide, prior to being matched with an employer, whether making the skill investment is worthwhile. An agent with cost c invests in skills ($e = 1$) if and only if the anticipated return from doing so exceeds this cost for the skill acquisition.

The newborn agents are also allowed to choose how they are perceived in the society. There are two types of affects that they can assume, either A or B: $i \in \{A, B\}$. They can choose how to present themselves either way incurring some cost. The relative cost of being perceived as A rather than B is $k \in R$. This can be positive or negative. If it is positive, he is naturally inclined to be perceived as B and should incur the cost k to be perceived as A. If it is negative, he is naturally inclined to be perceived as A and should incur the cost $-k$ to be perceived as B. The cumulative distribution function (CDF) of the cost is denoted by $H(k)$. We assume the symmetry of the distribution: $H(k) = 1 - H(-k)$. The population does not incline to one way or the other, implying that half of the population is naturally inclined toward A and the other half is naturally inclined toward B. An agent with cost k chooses to be perceived as A if and only if the incentive for electing the A-type rather than the B-type exceeds the relative cost for being perceived as A.

We state that there is no connection between the two exogenous cost variables, c and k. The economic ability of an individual and the natural affect orientation of an individual are distributed independently in the population.

For the wage setting mechanism, we adopt a statistical discrimination framework originally proposed in Coate and Loury(1993), which links the reputation of a group and the skill acquisition incentives for the group members. Employers cannot observe the skill level of a person, but they can observe the group to which the person belongs and a noisy signal $t \in [0, 1]$ that is generated out of the hiring process. The signal might be the result of the test, an interview by employers, internship, or on-the-job training. The distribution of the signal depends on whether the person has the skill. Let $F_1(t)\,[F_0(t)]$ be the probability that the signal does not exceed t, given that a worker is skilled (unskilled), and let $f_1(t)\,[f_0(t)]$ be the related density function. Define $\psi(t) \equiv f_1(t)/f_0(t)$, to be the likelihood ratio at t. We assume that $\psi(t)$ is a monotonically increasing function in t, which is defined as the Monotonic

Likelihood Ratio Property (MLRP). This property implies $F_1(t) < F_0(t)$ for any $t \in (0,1)$.[2] Thus, higher values of the signal are more likely if the worker is skilled, and for a given prior, the posterior likelihood that a worker is skilled is larger if his signal takes a higher value.

Employers start with a prior belief about the actual rate of skill acquisition of a group π. Let us define the function $f(\pi,t) \equiv \pi f_1(t) + (1-\pi) f_0(t)$, which indicates the distribution of the signal t of agents belonging to a group with the skill level π. The employers' posterior belief of the likelihood that an agent who presents the test score t is in fact skilled is achieved using the Bayes' rule: $\rho(\pi,t)(\equiv Pr[e=1|\pi,t]) = \frac{\pi f_1(t)}{f(\pi,t)}$. Imagine a simple economy in which the value of a skilled worker to employers is w and the value of an unskilled worker to employers is zero. The wage will be set to be proportional to the expected skill level. Then, the anticipated reward for an individual who belongs to a group with the believed skill acquisition rate of π and whose test score is realized as t is

$$\begin{aligned} W(\pi,t) &= w \cdot \rho(\pi,t) \\ &= w \cdot \frac{\pi f_1(t)}{\pi f_1(t) + (1-\pi) f_0(t)}. \end{aligned} \tag{1}$$

Given this framework, we can readily express the expected reward from acquiring a skill and that without acquiring a skill as follows:

$$V_e(\pi) = \int_0^1 f_e(t) W(\pi,t)\,\mathrm{d}t, \ \forall e \in \{0,1\}, \tag{2}$$

in which both $V_0'(\pi)$ and $V_1'(\pi)$ are positive for any π because the first derivatives are derived as[3]

$$V_0'(\pi) = \int_0^1 w f_1(t) f_0(t)^2 f(\pi,t)^{-2}\,\mathrm{d}t, \tag{3}$$

$$V_1'(\pi) = \int_0^1 w f_1(t)^2 f_0(t) f(\pi,t)^{-2}\,\mathrm{d}t. \tag{4}$$

Thus, both $V_0(\pi)$ and $V_1(\pi)$ are increasing functions of the believed skill acquisition rate π,

[2]Denote $\bar{t}$ which satisfies $\frac{f_1(\bar{t})}{f_0(\bar{t})} = 1$. For any $t \in (0,\bar{t})$, the following holds $F_1(t) - F_0(t) = \int_0^t f_1(x)(1 - \frac{f_0(x)}{f_1(x)})\,\mathrm{d}x < 0$. For any $t \in [\bar{t},1)$, the following holds $F_1(t) - F_0(t) = -\int_{\bar{t}}^1 f_1(x)(1 - \frac{f_0(x)}{f_1(x)})\,\mathrm{d}x < 0$.

[3]Using these derivatives, we can derive that$\lim_{\pi\to 0} V_0'(\pi) = w$ and $\lim_{\pi\to 1} V_1'(\pi) = w$. Because we know that $\lim_{\pi\to 0} R'(\pi) > 0$ and $\lim_{\pi\to 1} R'(\pi) < 0$ from equation (6) and MLRP property, the following must hold: $\lim_{\pi\to 0} V_1'(\pi) > w$ and $\lim_{\pi\to 1} V_0'(\pi) > w$.

as depicted in Panel A of Figure 1.

Workers' expected economic return from being skilled, which is denoted by $R(\pi)$, is equivalent to the difference between the expected reward from acquiring a skill and that without acquiring a skill: $R(\pi) \equiv V_1(\pi) - V_0(\pi)$. For instance, if the employers' prior belief is that none of the randomly selected members of a group has invested in skills ($\pi = 0$), the posterior belief about the likelihood that an agent of the group invested in a skill is zero ($\rho = 0$) regardless of the realization of the test score t, which means the anticipated reward for the agent belonging to the group is also zero regardless of the test score: $W(0,t) = 0, \forall t \in (0,1)$. Therefore, given $\pi = 0$, the expected wage is zero, regardless of whether he invests in the skill, implying that the expected economic return from being skilled is zero: $V_1(0) = V_0(0) = 0$ and $R(0) = 0$. In contrast, if employers are certain a priori that a randomly selected member from this population has invested in skill ($\pi = 1$), the posterior belief about the likelihood that an agent of this group invested in a skill is one ($\rho = 1$), given any test score realized, which means the anticipated reward for this agent is always w: $W(1,t) = w, \forall t \in (0,1)$. Therefore, regardless of whether he invests in a skill, the expected reward is fixed as w, implying that the expected economic return from being skilled is zero again: $V_1(1) = V_0(1) = w$ and $R(1) = 0$. Hence, we determine that both $R(0)$ and $R(1)$ are zero.

Using the above equations, the expected economic return from skill investment for an individual who belongs to a group with the believed investment rate of π is expressed as

$$R(\pi) = w\pi \int_0^1 \frac{(f_1(t) - f_0(t))f_1(t)}{f(\pi,t)}\,\mathrm{d}t, \tag{5}$$

which confirms that $R(0) = R(1) = 0$. The first and second derivatives of the return function can be directly seen as:

$$R'(\pi) = w \int_0^1 \frac{(f_1(t) - f_0(t))f_1(t)f_0(t)}{f(\pi,t)^2}\,\mathrm{d}t, \tag{6}$$

$$R''(\pi) = -2w \int_0^1 \frac{(f_1(t) - f_0(t))^2 f_1(t)f_0(t)}{f(\pi,t)^3}\,\mathrm{d}t. \tag{7}$$

Using MLRP property, we can derive that $\lim_{\pi\to 0} R'(\pi) > 0$ and $\lim_{\pi\to 1} R'(\pi) < 0$.[4] Because

[4] $\lim_{\pi\to 0} R'(\pi) = w\int_0^1 [f_1(t) - f_0(t)] \cdot \frac{f_1(t)}{f_0(t)}\,\mathrm{d}t = w\int_0^{\bar{t}} [f_1(t) - f_0(t)] \cdot \frac{f_1(t)}{f_0(t)}\,\mathrm{d}t + w\int_{\bar{t}}^1 [f_1(t) - f_0(t)] \cdot \frac{f_1(t)}{f_0(t)}\,\mathrm{d}t > w\int_0^{\bar{t}} [f_1(t) - f_0(t)]\,\mathrm{d}t + w\int_{\bar{t}}^1 [f_1(t) - f_0(t)]\,\mathrm{d}t = 0$, in which $\bar{t}$ satisfies $\frac{f_1(\bar{t})}{f_0(\bar{t})} = 0$. In the same way, we can indicate that $\lim_{\pi\to 1} R'(\pi) < 0$.

the second derivative of the return function is negative for any π, $R(\pi)$ is concave. The return is maximized at $\bar{\pi}$, which satisfies $R'(\bar{\pi}) = 0$. Panel B of Figure 1 illustrates how this return function $R(\pi)$, an agent's skill acquisition incentive, depends upon his group's collective reputation π.

Finally, a worker with skill acquisition cost c, who belongs to a group believed to be investing at rate π, has the anticipated net reward of $V_1(\pi) - c$ if he decides to be a skilled person and that of $V_0(\pi)$ if he decides not to be skilled. Thus, the anticipated net reward in the labor market for such a worker, $U(\pi, c)$, is summarized as

$$U(\pi, c) = \max\{V_1(\pi) - c,\ V_0(\pi)\}, \tag{8}$$

in which the function $U(\pi, c)$ is increasing in π for both $V_1'(\pi)$ and $V_0'(\pi)$ are positive for any π. We also know that the function is non-increasing in c given any specific π.

3 Phenotypic vs Endogenous Stereotyping Equilibria

In this section, we consider how the employers' belief about some group can be confirmed in an equilibrium involving rational behavior by the group members. We define both the Phenotypic and Endogenous Stereotyping Equilibria below.

3.1 Phenotypic Stereotyping Equilibria

We assume initially that society consists of exogenous, visibly distinct and equally endowed groups, the membership of which is immutable. Then, employers can discriminate among individuals based upon this observable 'phenotype'.

If employers anticipate that the probability that a randomly drawn individual from a population group i has invested in a skill is π_i, the return of the individual belonging to this group from investing in skill is $R(\pi_i)$. Then, the fraction of the group who will invest is $G(R(\pi_i))$, given the skill acquisition cost distribution $G(c)$. Thus, when a prior belief π_i satisfies $G(R(\pi_i)) = \pi_i$, such a belief about any group is self-confirming. Let us denote an equilibrium belief by $\hat{\pi} \in [0, 1] : \hat{\pi} = G(R(\hat{\pi}))$. The set of all such equilibrium beliefs is denoted by Ψ_{CL} (Coate an Loury, 1993). We call such outcomes "Phenotypic Stereotyping Equilibria (PSE)." An example of such equilibria is described in Panel B of Figure 1, in which

$R(\pi)$ is concave and $G(c)$ is S-shaped.

Multiple equilibria create the possibility of phenotypic stereotyping wherein exogenously and visibly distinct groups fare unequally in the equilibrium. When group identity is exogenous and Ψ_{CL} contains multiple elements, unequal reputations of the groups can be sustained in equilibrium despite the groups being equally well endowed (i.e., having the same $G(c)$). In this case, inequality of collective reputation between the exogenous groups in equilibrium is due to the feedback between the group reputation and individual skill investment activities. The individuals in a group with a better collective reputation have a greater incentive to invest in skills, and with their greater skill investment rate, the group maintains a better collective reputation (and vice versa). The stereotypic beliefs about the exogenous groups are self-confirmed.

3.2 Endogenous Stereotyping Equilibria

When group membership is exogenous, the skill acquisition cost distribution, namely, $G(c)$, is the same for any group. However, when group membership is endogenous, this is no longer true. Given that the collective reputations of the two affective groups differ and the expected economic return from being skilled, namely, $R(\pi)$, is greater for the favored group, the favored group not only faces greater human capital investment incentives, but it also consists disproportionately of lower skill acquisition cost types, who gain most from joining this favored group. Thereby, it will cause human capital cost distributions between groups to endogenously diverge, reinforcing incentive-feedbacks. The following displays the equilibria that are the result of this phenomenon.

Consider a society in which workers can choose a perceived group membership, A or B, though at some cost k (either positive or negative) of affecting identity "A". Let a and b be employer beliefs about human capital investment rates in affective groups A and B. $U(a, c)$ ($U(b, c)$) is the anticipated net reward in the labor market for an agent who is perceived as a member of group A (group B) and whose skill acquisition cost is given as c. Let us define a function $\Delta U(a, b; c)$ as the payoff difference between an A-type worker and a B-type worker given their skill acquisition cost level c: $\Delta U(a, b; c) \equiv U(a, c) - U(b, c)$. This indicates the incentive for electing type-A rather than B-type for an agent whose skill acquisition cost is c. Symmetrically, $\Delta U(b, a; c) \equiv U(b, c) - U(a, c)$, indicating the incentive for electing type-B

rather than type-A. When $a > (<) b$, $\Delta U(a, b; c)$ is positive (negative) because $U(\pi, c)$ is increasing in π. Note that $\Delta U(a, b; c) = -\Delta U(b, a; c)$ and $\Delta U(a, b; c) = 0$ when $a = b$.

Agents are endowed with two cost parameters: cost of skill (c) and cost of affect "A" (k). An agent with the cost set (c, k) elects to be an A-type worker if and only if $\Delta U(a, b; c) \geq k$. Otherwise, he elects to be a B-type worker. When the employers' prior beliefs about the actual skill acquisition rates of the groups are identical (i.e., $a = b$ and $\Delta U(a, b; c) = 0$), an agent with positive (negative) k naturally elects to be a type-B (type-A) regardless of skill acquisition cost level c. Because c and k are independently distributed, the fraction of workers who elect to be A-type is $H(\Delta U(a, b; c))$ among the population segment with skill acquisition cost level c. Thus, among the whole population, the fraction of agents who elect to be A-type is given by using the two cumulative distribution functions $H(k)$ and $G(c)$,

$$\Sigma^A \equiv \int_0^\infty H(\Delta U(a, b; c))\, \mathrm{d}G(c). \tag{9}$$

Among the agents who will elect to be A-type, the higher capability population whose skill acquisition cost is not greater than the incentives for skill investment ($c \leq R(a)$) will elect to be skilled. Then, the fraction of workers who elect to be A-type and become skilled is given by

$$\sigma^A \equiv \int_0^{R(a)} H(\Delta U(a, b; c))\, \mathrm{d}G(c). \tag{10}$$

Among the population whose skill acquisition cost level is c, the fraction of agents who elect to be B-type is $1 - H(\Delta U(a, b; c))$, which is equivalent to $H(\Delta U(b, a; c))$ by the symmetry assumption of $H(k) = 1 - H(-k)$. Thus, among the total population, the fraction of agents who elect to be B-type is given by

$$\Sigma^B \equiv \int_0^\infty H(\Delta U(b, a; c))\, \mathrm{d}G(c). \tag{11}$$

Among the agents who will elect to be B-type, the population whose skill acquisition cost is not greater than the incentives for skill investment ($c \leq R(b)$) will elect to be skilled. Consequently, the fraction of workers who elect to be B-type and become skilled is given by

$$\sigma^B \equiv \int_0^{R(b)} H(\Delta U(b, a; c))\, \mathrm{d}G(c). \tag{12}$$

Therefore, given the employer belief about human capital investment rates (a, b), the actual investment rates for the endogenously constructed groups A and B are denoted by $\phi(a; b)$ $(= \sigma^A/\Sigma^A)$ and $\phi(b; a)$ $(= \sigma^B/\Sigma^B)$ for each, where the function $\phi(x; y)$ is defined as follows:

$$\phi(x; y) \equiv \frac{\int_0^{R(x)} H(\Delta U(x, y; c))\, \mathrm{d}G(c)}{\int_0^{\infty} H(\Delta U(x, y; c))\, \mathrm{d}G(c)}, \tag{13}$$

in which $\phi(x; x) = G(R(x))$.

An equilibrium in this society with endogenous group membership is defined as a pair of investment rates for the endogenously constructed groups $(a^*, b^*) \in [0, 1]^2$ such that $a^* = \phi(a^*; b^*)$ and $b^* = \phi(b^*; a^*)$. We call such outcomes "Endogenous Stereotyping Equilibria (ESE)," and the set of all such equilibria is denoted by Ω_{KL}. Note that every PSE corresponds to trivial ESE for which differences in affect are uninformative: $(\hat{x}, \hat{x}) \in \Omega_{KL}$ if $\hat{x} \in \Psi_{CL}$ because $\phi(\hat{x}, \hat{x}) = G(R(\hat{\pi})) = \hat{x}$. Our main interest is with the non-trivial ESE, in which $a^* \neq b^*$.

4 Properties of Identity Choice Behaviors

In this section, we examine the key properties of the identity choice behaviors. The expected payoff difference between an A-type worker and a B-type worker in the labor market, $\Delta U(a, b; c)$, can be expressed by using the equation (8),

$$\Delta U(a, b; c) = \max\{R(a), c\} - \max\{R(b), c\} + V_0(a) - V_0(b). \tag{14}$$

Then, we have the following lemma concerning $\Delta U(a, b; c)$:

Lemma 1. *For any* $c \leq \min\{R(a), R(b)\}$, $\Delta U(a, b; c) = V_1(a) - V_1(b)$. *For any* $c \geq \max\{R(a), R(b)\}$, $\Delta U(a, b; c) = V_0(a) - V_0(b)$. *For any* c *such that* $\min\{R(a), R(b)\} < c < \max\{R(a), R(b)\}$, *we have*

$$\Delta U(a, b; c) = \begin{cases} V_1(a) - V_0(b) - c & \text{if } R(a) \geq R(b), \\ V_0(a) - V_1(b) + c & \text{if } R(a) < R(b). \end{cases} \tag{15}$$

The above lemma is summarized in Figure 2. The four panels of the figure describe

$\Delta U(a, b; c)$ with respect to skill acquisition cost level c for the following distinct cases: $a > b$ and $R(a) > R(b)$ (Panel A), $a > b$ but $R(a) < R(b)$ (Panel B), $a < b$ but $R(a) > R(b)$ (Panel C) and $a < b$ and $R(a) < R(b)$ (Panel D). Among them, the first case is the most relevant situation in the subsequent arguments: The positive complementarity between collective reputation and the skill investment incentives given a privileged group A and a disadvantaged group B, which is also emphasized in Panel A of Figure 1 displaying $R(a) > R(b)$ given $a > b$. Note that the relationship must entail that $V_1(a) - V_1(b) > V_0(a) - V_0(b)$.

Form the above lemma, we achieve two important propositions resulting from the identity choice behaviors of economic agents. First, it is directly seen that $\Delta U(a, b; c) > (<) \, 0$ for any given cost level c if and only if $a > (<) \, b$, as displayed in Panels A and B (Panels C and D), which is summarized in the following proposition.

Proposition 1. *When employers have different beliefs about two affective groups($a \neq b$), the fraction of workers who adopt the 'affect' corresponding to the favored employers' belief is greater than that of workers who adopt the 'affect' with the less favored employers' belief: $\Sigma^A > (<) \, \Sigma^B$ if $a > (<) \, b$.*

That is, in the current setting with a symmetric cost distribution of $h(k)$, more than half of workers adopt the 'affect' that corresponds to the more favorable employers' belief: $\Sigma^A > .5$ and $\Sigma^B < .5$ if $a > b$. Additionally, Lemma 1 indicates that $\Delta U(a, b; c)$ is non-increasing with respect to c whenever $R(a) > R(b)$, as depicted in Panels A and C, which implies $\phi(a; b) > G(R(a))$ and $\phi(b; a) < G(R(b))$. In a symmetric way, $\Delta U(a, b; c)$ is non-decreasing with respect to c whenever $R(a) < R(b)$, as depicted in Panels B and D, which implies $\phi(a; b) < G(R(a))$ and $\phi(b; a) > G(R(b))$. However, when $R(a) = R(b)$ even with $a \neq b$, $\Delta U(a, b; c)$ is constant with respect to c, and we have $\phi(a; b) = \phi(b; a) = G(R(a)) = G(R(b))$. These properties are summarized by the following proposition.

Proposition 2. *The disproportionately more talented workers, whose human capital investment costs (c) are relatively lower, choose the 'affect' that corresponds to the greater return to human capital investment given $R(i) > R(j)$, $\phi(i; j) > G(R(i))$ and $\phi(j; i) < G(R(j))$ for each combination $(i, j) \in \{(a, b), (b, a)\}$.*

For further analysis, we need concrete information about the skill investment rates $\phi(a; b)$ resulting from the identity choice behaviors. The overall shape of $\phi(a; b)$ with respect to

a given a fixed level of b is displayed in Figure 3 for three different levels of b below $\bar{\pi}$, $b_1 < b_2 < b_3 < \bar{\pi}$, together with the benchmark curve $\phi(a;a)(= G(R(a)))$. Note that for any b except for $\bar{\pi}$, we can find $b'(\neq b)$ such that $R(b) = R(b')$. As discussed above, the following should hold for the combination (b, b'): $\phi(b;b) = \phi(b';b) = G(R(b)) = G(R(b'))$. Therefore, we find that the dotted $\phi(a;b)$ curve intercepts the solid $\phi(a;a)(= G(R(a)))$ curve at $a = b$ and $a = b'$ in the figure. WLOG, we impose the following smoothing condition for the relative positions of the $\phi(a;b)$ curves.

Condition 1. *For any $\tilde{b}'$ between a and b, $\phi(a;\tilde{b}')$ is also between $\phi(a;a)$ and $\phi(a;b)$:*

$$min\{\phi(a;a), \phi(a;b)\} < \phi(a;\tilde{b}') < max\{\phi(a;a), \phi(a;b)\}, \forall \tilde{b}' \in (a, b). \tag{16}$$

This condition implies that for any specific b_1 and b_2 such that $b_1 < b_2 < \bar{\pi}$, the $\phi(a;b_1)$ curve is placed above the $\phi(a;b_2)$ curve: $\phi(a;b_1) > \phi(a;b_2)$, $\forall a \in (0,1)$, as also displayed in Figure 3. Additionally, for any specific b_1 and b_2 such that $\bar{\pi} < b_1 < b_2$, the $\phi(a;b_2)$ curve is placed above the $\phi(a;b_1)$ curve: $\phi(a;b_2) > \phi(a;b_1)$, $\forall a \in (0,1)$. The following lemma helps us understand the curvature of the $\phi(a;b)$ curve when it crosses over the $\phi(a;a)$ curve:

Lemma 2. *The slope of the $\phi(a;b)$ curve at the point where it crosses over the $\phi(a;a)$ curve is*

$$\frac{\partial \phi(a;b)}{\partial a}\Big|_{a=b} \approx g(R(b))R'(b) + 2H'(0)R'(b)G(R(b)) \cdot [1 - G(R(b))]. \tag{17}$$

Proof. See the proof in the appendix. ■

The above lemma implies that the slope of $\phi(a;b)$ at the crossing point is positive (negative) whenever $R'(b)$ is positive (negative). Additionally, the slope of $\phi(a;b)$ at the crossing point is greater (smaller) than the slope of $\phi(a;a)(= g(R(b))R'(b))$ whenever $R'(b)$ is positive (negative), which means the slope of $\phi(a;b)$ is "steeper" than $\phi(a;a)$ at such crossing point.

5 Characteristics of Endogenous Stereotyping Equilibria

Now, we are ready to examine both the existence and the stability of Endogenous Stereotyping Equilibria. For this purpose, we introduce a correspondence $\Gamma(y)$: $\Gamma(y) = \{x \mid x = \phi(x;y)\}$. By definition, the correspondence indicates interceptions between the $\phi(x;y)$ curve and 45 degree

line, at which a group's actual skill investment rate $\phi(x; y)$ becomes equal to the employers' prior belief about the group's skill level x, given the employers' prior belief about the other group's skill level y. For example, given b_1, the $\phi(a; b_1)$ curve intercepts the 45-degree line three times in Figure 3. The three crossing points marked with tiny triangles represent the correspondence $\Gamma(b_1)$. First, note that any $\hat{\pi} \in \Psi_{CL}$ satisfies $\hat{\pi} \in \Gamma(\hat{\pi})$ and any $\hat{\pi} \in \Gamma(\hat{\pi})$ satisfies $\hat{\pi} \in \Psi_{CL}$. Thus, the set of phenotypic stereotyping equilibria (PSE) is represented as follows using the correspondence: $\Psi_{CL} = \{x \,|\, x \in \Gamma(x)\}$. On the other hand, the set of endogenous stereotyping equilibria (ESE) is expressed as $\Omega_{KL} = \{(x, y) \,|\, x \in \Gamma(y) \text{ and } y \in \Gamma(x)\}$ because an ESE is defined as a pair (x, y) that satisfies both $x = \phi(x; y)$ and $y = \phi(y; x)$.

In the following discussions, we consider the case with multiple PSE, in which, WLOG, we assume that there are three equilibria: π_l, π_m and π_h, with the ordering of $\pi_l < \pi_m < \pi_h$. For the concise presentation of our key arguments, we further impose that the three equilibria are placed below $\bar{\pi}$: $\pi_h < \bar{\pi}$. However, readers will find that the main results do not change for the other case with $\pi_h > \bar{\pi}$, which is not presented in this manuscript.[5]

5.1 Existence of Endogenous Stereotyping Equilibria

When there are three unique values in a correspondence $\Gamma(y)$, let us denote the greatest, the middle and the smallest one of them by $\Gamma(y)^h$, $\Gamma(y)^m$ and $\Gamma(y)^l$ for each in the $(y, \Gamma(y))$ plane, as displayed in Figure 4. When the correspondence $\Gamma(y)$ contains just one value, the value is denoted by $\Gamma(y)^i$ as it is connected along the correspondence curve to nearby $\Gamma(y)^i$ for $i \in \{h, m, l\}$ in the plane. (Refer to Figure 4.) Under the smoothing condition in equation (16), we know that $\Gamma(y)^h$ and $\Gamma(y)^l$ decline and $\Gamma(y)^m$ increases, as y increases over the range $(0, \bar{\pi})$ and as y decreases over the range $(\bar{\pi}, 1)$. Under the condition of $\pi_h < \bar{\pi}$, we can easily infer that $\Gamma(\bar{\pi})$ contains a unique value from Figure 3, which will be denoted by $\Gamma(\bar{\pi})^l$ as it is connected to nearby $\Gamma(y)^l$. The curvature of the correspondence $\Gamma(y)$ is summarized in the following lemma:

Lemma 3. *For any y below $\bar{\pi}$, $\Gamma(y)^h$ and $\Gamma(y)^l$ decrease in y and $\Gamma(y)^m$ increases in y, while $\Gamma(y)^h$ and $\Gamma(y)^l$ increase in y and $\Gamma(y)^m$ decreases in y for any y above $\bar{\pi}$.*

This lemma implies that $\min[\Gamma(y)^l] = \Gamma(\bar{\pi})^l$ and $\arg\min[\Gamma(y)^l] = \bar{\pi}$. Additionally, we have $\pi_h < \Gamma(0)^h < 1$ and $\pi_h < \Gamma(1)^h < 1$. Based on the above findings, the two correspondences

[5]The analysis with $\pi_h > \bar{\pi}$, which is not included in this manuscript, can be delivered upon request.

$\Gamma(a)$, which is a set $\{b|b = \phi(b; a)\}$, and $\Gamma(b)$, which is a set $\{a|a = \phi(a; b)\}$, are overlapped in Figure 4. Using the local linearization process, we can calculate the slope of correspondence curve at each trivial ESE, which satisfies $\hat{\pi} \in \Gamma(\hat{\pi})$.

Lemma 4. *The slope of the correspondence curve at a trivial ESE $(\hat{x}, \hat{x})$, which is denoted by $\Gamma'(\hat{x})$, is approximated by*

$$\Gamma'(\hat{x}) \approx \frac{2H'(0)R'(\hat{x})\hat{x}(1-\hat{x})}{g(R(\hat{x}))R'(\hat{x}) - 1 + 2H'(0)R'(\hat{x})\hat{x}(1-\hat{x})}. \tag{18}$$

Proof. See the proof in the appendix. ■

The examples for the positive $\Gamma'(\hat{x})$ and the negative $\Gamma'(\hat{x})$ are depicted in Appendix Figure 1. Using the above lemma, we can describe the correspondence curves more accurately. Before we address the results, we acknowledge the followings three facts: (1) $R'(\hat{x})$ is positive for any PSE $\hat{x}$ because we assume $\pi_h < \bar{\pi}$; (2) The slope of the $\phi(a; a)$ curve at $a = \pi_m$ is greater than one: $g(R(\pi_m))R'(\pi_m) > 1$; (3) The slope of the curve at $a = \pi_h$ (or π_l) is smaller than one: $0 < g(R(\hat{x}))R'(\hat{x}) < 1$, $\forall \hat{\pi} \in \{\pi_h, \pi_l\}$. You may refer to Figure 4 for these primary facts. The slope of the correspondence curve, $\Gamma'(\hat{x})$, can be presented depending on the density of the identity choice cost distribution around zero, $H'(0)$. The following lemma summarizes the results.

Lemma 5. *The slope of the correspondence curve at a trivial ESE (π_m, π_m) always satisfies $0 < \Gamma'(\pi_m) < 1$. The slope of the correspondence curve at a trivial ESE, either (π_h, π_h) or (π_l, π_l), depends on the density of the identity choice cost distribution around zero, $H'(0)$:*

$$\begin{cases} -1 < \Gamma'(\hat{x}) < 0 & \textit{if} \quad H'(0) < \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})} \\ \Gamma'(\hat{x}) < -1 & \textit{if} \quad \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})} < H'(0) < \frac{1-g(R(\hat{x}))R'(\hat{x})}{2R'(\hat{x})\hat{x}(1-\hat{x})}, \quad \forall \hat{x} \in \{\pi_h, \pi_l\}. \\ \Gamma'(\hat{x}) > 1 & \textit{if} \quad H'(0) > \frac{1-g(R(\hat{x}))R'(\hat{x})}{2R'(\hat{x})\hat{x}(1-\hat{x})} \end{cases} \tag{19}$$

Proof. Using $g(R(\pi_m))R'(\pi_m) > 1$ and $0 < g(R(\hat{x}))R'(\hat{x}) < 1$, $\forall \hat{\pi} \in \{\pi_h, \pi_l\}$, we can directly derive the results from Lemma 4. ■

The lemma implies that when the sensitivity of identity choice activities represented by $H'(0)$ is sufficiently high that it is greater than $\frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})}$, the absolute value of the slope of correspondence curve $|\Gamma'(\hat{x})|$ at a trivial ESE $(\hat{x}, \hat{x})$, $\forall \hat{x} \in \{\pi_h, \pi_l\}$ is greater than one.

Furthermore, at least two non-trivial ESE exist as far as multiple PSE exist. This can be proved directly using the overlapped shapes of $\Gamma(a)$ and $\Gamma(b)$ in the (a, b) coordination plane:

Proposition 3. *Given multiple PSE (π_l, π_m and π_h), there always exist at least two non-trivial ESE.*

Proof. See the proof in the appendix. ■

Whether there are more than two ESE depends on the curvature of $\Gamma(a)$ and $\Gamma(b)$ around trivial ESE $(\hat{x}, \hat{x})$. The slope of the correspondence curve at a trivial ESE, $\Gamma'(\hat{x})$, determines the exact number of non-trivial ESE. WLOG, the condition $|\Gamma'(\hat{x})| < 1$ for each $\hat{x} \in \{\pi_h, \pi_l\}$ generates two non-trivial ESE around a trivial ESE $(\hat{x}, \hat{x})$, while the condition $|\Gamma'(\hat{x})| > 1$ for each $\hat{x} \in \{\pi_h, \pi_l\}$ does not generate such additional non-trivial ESE around a trivial ESE $(\hat{x}, \hat{x})$. Refer to Panel A of Figure 4 for the case with the condition $|\Gamma'(\hat{x})| < 1$ for each $\hat{x} \in \{\pi_h, \pi_l\}$, in which the total six non-trivial ESE are generated and Panels B and C of the figure for the case with the condition $|\Gamma'(\hat{x})| > 1$ for each $\hat{x} \in \{\pi_h, \pi_l\}$, in which only two non-trivial ESE are generated.

Proposition 4. *WLOG, the total number of non-trivial ESE is six when both $|\Gamma'(\pi_h)| < 1$ and $|\Gamma'(\pi_l)| < 1$, and it is only two when both $|\Gamma'(\pi_h)| > 1$ and $|\Gamma'(\pi_l)| > 1$.*

Let us call the two non-trivial ESE that exist regardless of the curvatures of the correspondences $\Gamma(a)$ and $\Gamma(b)$ "Persistent ESE" and denote them (π_L^*, π_H^*) and (π_H^*, π_L^*).

Theorem 1. *Given multiple PSE (π_l, π_m and π_h), there always exist two "Persistent ESE", (π_L^*, π_H^*) and (π_H^*, π_L^*), which satisfy $\pi_L^* < \pi_l < \pi_h < \pi_H^*$.*

Proof. See the proof in the appendix. ■

The theorem implies that the inequality between endogenous groups in some non-trivial ESE can be greater than the inequality between exogenous groups in any PSE.

5.2 Stability of Endogenous Stereotyping Equilibria

Consider an intergenerational population structure. Every period, the randomly chosen α fraction of the workers die and the same number of agents are newly born. The newborn agents incur the cost c of skill achievement, and the cost k to choose the affect A: k can be positive or negative. Each newborn agent with his cost set (c, k) decides whether to invest for

skills and which 'affect' to choose among A and B in the early days of his life. After the days of education and affect adoption, newborns join the labor market and receive wages set by employers. We assume that employers set the newborn lifetime wage $W(\pi, t)$ proportional to the estimated skill level $\rho(\pi, t)$: $W(\pi_j, t) = w \cdot \rho(\pi_j, t)$, given $\rho(\pi_j, t) = \pi_j f_1(t)/f(\pi_j, t)$, for the entering newborns with the perceived identity $j \in \{A, B\}$ and the noisy signal t. Employers compare the actual skill acquisition rate of the newborns who adopt the affect j, $\phi(\pi_j; \pi_{-j})$, and the believed overall skill rate π_j of the workers belonging to identity group j to update their prior belief π_j.[6] Therefore, we have the following dynamics:

$$\dot{\pi}_j > (<)\, 0 \Leftrightarrow \phi(\pi_j; \pi_{-j}) > (<)\, \pi_j. \tag{20}$$

At the bottom of Figure 3, we present the law of motions of a given an arbitrary b_1: $\dot{a} > 0$ for any $a \in (0, \Gamma(b_1)^l)$ and any $a \in (\Gamma(b_1)^m, \Gamma(b_1)^h)$, and $\dot{a} < 0$ for any $a \in (\Gamma(b_1)^l, \Gamma(b_1)^m)$ and any $a \in (\Gamma(b_1)^h, 1)$. Therefore, the direction arrows of $\dot{a}$ are upward between $\Gamma(b)^m$ and $\Gamma(b)^h$ and below $\Gamma(b)^l$ in the (b, a) coordination plane and downward between $\Gamma(b)^l$ and $\Gamma(b)^m$ and above $\Gamma(b)^h$, as displayed in Figure 4. In a symmetric way, the direction arrows of $\dot{b}$ are rightward between $\Gamma(a)^m$ and $\Gamma(a)^h$ and at the left-hand side of $\Gamma(a)^l$ in the (b, a) coordination plane and leftward between $\Gamma(a)^l$ and $\Gamma(a)^m$ and at the right-hand side of $\Gamma(a)^h$. From the described direction arrows, we can infer the following theorem.

Theorem 2. *Given multiple PSE (π_l, π_m and π_h), two "Persistent ESE", (π_L^*, π_H^*) and (π_H^*, π_L^*), are stable and all other non-trivial ESE are unstable.*

The theorem together with Theorem 1 implies that the inequality that can be generated between endogenously formed groups should be greater than the inequality between exogenous groups in any PSE in the long run because all non-trivial ESE are unstable except for "Persistent ESE." Using the direction arrows, we can easily confirm that the middle trivial ESE (π_m, π_m) is always unstable. Other trivial ESE, (π_h, π_h) and (π_l, π_l), are stable if $|\Gamma'(\hat{x})| \leq 1$ and unstable if $|\Gamma'(\hat{x})| > 1$. Using Lemma 5, we determine an additional lemma concerning the trivial ESE.

[6] We assume that employers have correct information about the actual skill acquisition rate of the newborns belonging to each identity group j. This information $(\phi(\pi_j; \pi_{-j}))$ can be deduced from the equality between the predicted average productivity of the newborn cohort adopting affect j and presenting the test score t, $w \cdot \frac{\phi(\pi_j;\pi_{-j}) f_1(t)}{\phi(\pi_j;\pi_{-j}) f_1(t) + (1-\phi(\pi_j;\pi_{-j})) f_0(t)}$, and the observed actual productivity of that cohort who presented score t in the screening process.

Lemma 6. *Given multiple PSE (π_l, π_m and π_h), the trivial ESE (π_m, π_m) is always unstable, while the other trivial ESE, (π_h, π_h) or (π_l, π_l), is unstable if and only if it satisfies $H'(0) > \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})}$, for $\hat{x} \in \{\pi_h, \pi_l\}$.*

Therefore, when the society consists of a sufficiently large fraction of newborns whose identity choice cost is very low (i.e., $H'(0)$ is sufficiently large), the balanced skill rates between two identity groups, (π_h, π_h) or (π_l, π_l), cannot be sustainable due to the incentives for the talented members to choose the "affect" associated with the better collective reputation. This lemma implies the following worthwhile result:

Proposition 5. *While "Persistent ESE", (π_L^*, π_H^*) and (π_H^*, π_L^*), are always stable, all other ESE are unstable if and only if $H'(0) > \frac{1-g(R(\hat{x}))R'(\hat{x})}{4R'(\hat{x})\hat{x}(1-\hat{x})}, \forall \hat{x} \in \{\pi_h, \pi_l\}$.*

When the sensitivity of identity choice activities is sufficiently high, the skill composition of the society must converge to a "Persistent ESE" in the long run, in which inequality between endogenously formed identity groups is greater than between exogenously given groups in any PSE: $|\pi_H^* - \pi_L^*| > |\pi_i - \pi_j|, \forall i, j \in \{l, m, h\}$. The egalitarian policies between endogenously constructed social groups may not be successful when identity manipulation is easier to undertake because any equal society cannot be sustainable as the more talented individuals adopt the more highly regarded group's identity disproportionately, letting the between-group difference diverge over time.

6 Applications and Welfare Properties

The theory developed in this paper can be applied to many situations in which identity choice and group stereotypes operate in tandem. In this section, we discuss the welfare effects resulting from the following two human behaviors associated with the manipulation of the perceived identity. The first is the selective out-migration from a stigmatized group to a better-regarded group, which is often referred to as "passing" activities. The second is the production of the indices of differentiation by better off members of the negatively stereotyped population. We call this behavior "partial passing", following the definition in Loury (2002) of the strategies of social identity manipulation used by racially marked people to inhibit being stereotyped.

6.1 Selective Out-migration ("Passing")

The selective out-migration from a stigmatized group to a better regarded group occurs when the return for "passing" such as better treatment in the labor market outweighs its cost such as losing ties to ones' own kind, learning unfamiliar customs and adopting a new culture. To examine the welfare effects of the passing behaviors, both the returns and the costs should be included in the welfare function.

Consider two social groups, a privileged group (A) and a stigmatized group (B). According to Propositions 1 and 2, the incentive for individuals to join group A is greater than to join the stigmatized group B for any ability type (c), and the selective out-migration occurs as the more talented individuals are drawn to the favored group A disproportionately. According to Theorem 2, when the selective out-migration occurs, the only stable (non-trivial) equilibrium is a "Persistent ESE," in which the groups' collective reputations are self-confirmed at π_H^* and π_L^* for each.

The welfare effects of the passing behavior can be examined by comparing the welfare at this stable equilibrium to the welfare at the benchmark economy in which the choice of the perceived identity is not allowed: An agent with positive k belongs to the stigmatized group, while an agent with negative k belongs to the privileged group. Under the absence of the endogeneity of identity choice, each group's collective reputation is self-confirmed at one of the stable PSE (either π_h or π_l).[7] The collective reputations in equilibrium are π_h for the privileged group and π_l for the stigmatized group. Refer to Panel A of Figure 5 for this benchmark economy without the passing activities.

Now, let us clarify who benefits and who suffers from the existence of passing activities. The total population in the "Persistent ESE" can be classified into three population aggregates according to their identity manipulation incentives: "passers" who give up their natural orientation (type-B) to be perceived as type A ($0 < k < \Delta U(\pi_H^*, \pi_L^*; c)$), "non-passers" who maintain their natural orientation (type-B) although stigmatized in the marketplace ($k \geq \Delta U(\pi_H^*, \pi_L^*; c)$) and "the advantaged" who keep their privileged type-A membership ($k \leq 0$).

We need to examine the welfare change of each population aggregate between the "Persistent ESE" (π_L^*, π_H^*) and its PSE benchmark economy (π_l, π_h), using the condition $\pi_L^* < \pi_l <$

[7] Note that π_m is not stable in the sense that the group's overall skill acquisition rate $G(R(\pi))$ diverges away from π_m with any little perturbation.

$\pi_h < \pi_H^*$ according to Theorem 1. Remember that the function $U(\pi, c)$ indicates the anticipated net reward in the labor market for an individual who belongs to a group believed to be investing at rate π and whose skill acquisition cost level is c. Because the function is monotonically increasing in π, we can infer that the non-passers suffer from the allowed out-migration activities as much as $U(\pi_l, c) - U(\pi_L^*, c)$, while the advantaged benefit from such activities as much as $U(\pi_H^*, c) - U(\pi_h, c)$. Interestingly, not all of passers benefit from the existence of out-migrations. A passer' anticipated net reward changes as much as $U(\pi_H^*, c) - U(\pi_l, c) - k$ between the two distinct economies. Only those whose identity manipulation cost is sufficiently small that it is less than some threshold $\tilde{k}(c)$ benefit, while those whose identity manipulation cost is above the threshold suffer, in which $\tilde{k}(c) \equiv U(\pi_H^*, c) - U(\pi_l, c)$. Refer to Figure 6 for the three population aggregates and their welfare changes.

Denoting $\Delta W_{c,k}$ as the welfare change for an individual with two specific cost levels (c and k) from the PSE benchmark economy (π_l, π_h) to the "Persistent ESE" (π_L^*, π_H^*), we have the following proposition that summarizes the results:

Proposition 6. *Comparing an unequal stable economy with the passing activities (π_L^*, π_H^*) with the unequal stable economy without such activities (π_l, π_h), "Non-passers ($k > \Delta U(\pi_H^*, \pi_L^*; c)$" suffer from the selective out-migration, while the advantaged population ($k < 0$) benefits. Whether the welfare of a "passer" improves depends on the level of one's identity manipulation cost k. The welfare change of each population aggregate (non-passer, passer or the advantaged) between the two distinct economies is summarized as*

$$\Delta W_{c,k} = \begin{cases} U(\pi_L^*, c) - U(\pi_l, c), \forall k \in (\Delta U(\pi_H^*, \pi_L^*; c), \infty) & : \textit{negative; "non-passers"} \\ U(\pi_H^*, c) - U(\pi_l, c) - k, \forall k \in (0, \Delta U(\pi_H^*, \pi_L^*; c)) & : \textit{positive iff } k < \tilde{k}(c)\textit{; "passers"} \\ U(\pi_H^*, c) - U(\pi_h, c), \forall k \in (-\infty, 0) & : \textit{positive; "the advantaged"} \end{cases}$$

It is notable that we can always find the threshold $\tilde{k}(c)$ that satisfies $0 < \tilde{k}(c) < \Delta U(\pi_H^*, \pi_L^*; c)$ for any specific level of c. Thus, we achieve the following welfare property that denies the possibility of *parato improvement* from the passing activities.

Theorem 3. *Consider a privileged group A and a stigmatized group B. The self-confirmed skill acquisition rates are denoted by π_H^* and π_L^* with the existence of the passing activities and denoted by π_h and π_l with the absence of such activities, in which $\pi_L^* < \pi_l < \pi_h < \pi_H^*$ holds.*

Then, there always exists a threshold $\tilde{k}(c)$ $(\equiv U(\pi_H^*, c) - U(\pi_l, c))$ *between* 0 *and* $\Delta U(\pi_H^*, \pi_L^*; c)$*, so that the individuals whose identity manipulation cost is above the threshold suffer, while those whose identity manipulation cost is below the threshold benefit due to the existence of the passing activities.*

Second, let us examine the conditions under which the selective out-migration may improve the social efficiency. In the given theoretical framework, the employers' expected payoffs are always zero because they pay exact wages to workers according to their expected productivity. The social efficiency is thus determined by the welfare changes of workers between the two distinct economies. Using Proposition 6, we can compute the societal efficiency gain (ΔW_{total}) by the double integrations of $\Delta W_{c,k}$:[8]

$$\begin{aligned}
\Delta W_{total} &= \int_0^\infty \Big[\int_{\Delta U}^\infty [U(\pi_L^*, c) - U(\pi_l, c)]\, \mathrm{d}H(k) + \int_0^{\Delta U} [U(\pi_H^*, c) - U(\pi_l, c) - k]\, \mathrm{d}H(k) \\
&\quad + \int_{-\infty}^0 [U(\pi_H^*, c) - U(\pi_h, c)]\, \mathrm{d}H(k) \Big]\, \mathrm{d}G(c), \text{ where } \Delta U \equiv \Delta U(\pi_H^*, \pi_L^*; c) \\
&= \underbrace{0.5 \int_0^\infty [U(\pi_H^*, c) - U(\pi_h, c)]\, \mathrm{d}G(c)}_{\text{“positive reputational externality”}} - \underbrace{0.5 \int_0^\infty [U(\pi_l, c) - U(\pi_L^*, c)]\, \mathrm{d}G(c)}_{\text{“negative reputational externality”}} \\
&\quad + \underbrace{\int_0^\infty \int_0^{\Delta U} [\Delta U - k]\, \mathrm{d}H(k)\, \mathrm{d}G(c)}_{\text{“passing premium”}}, \text{ using the symmetry of } H(k).
\end{aligned}$$

The change from the PSE benchmark economy (π_l, π_h) to the "passing" equilibrium (π_L^*, π_H^*) generates the positive reputational externality for the population aggregate whose natural orientation is type A and the negative reputational externality for the population aggregate whose natural orientation is type B. The first term in the above equation indicates the size of positive reputational externality, while the second term indicates the size of negative reputational externality. The third term plays a significant role in the determination of the positive efficiency gain, which reflects the passing premium for the passers who choose to elect type A although their natural orientation is type B. The societal positive efficiency gain is achieved only when the passing premium is sufficiently great that it is bigger than the net loss in terms of the reputational externalities - the size of the negative reputational externailty

[8] Use the following decomposition: $\int_0^{\Delta U} [U(\pi_H^*, c) - U(\pi_l, c) - k]\, \mathrm{d}H(k) = \int_0^{\Delta U} [U(\pi_L^*, c) - U(\pi_l, c)]\, \mathrm{d}H(k) + \int_0^{\Delta U} [U(\pi_H^*, c) - U(\pi_L^*, c) - k]\, \mathrm{d}H(k) = \int_0^{\Delta U} [U(\pi_L^*, c) - U(\pi_l, c)]\, \mathrm{d}H(k) + \int_0^{\Delta U} [\Delta U - k]\, \mathrm{d}H(k)$.

minus the size of the positive reputational externality. The following summarizes the societal efficiency gain.

Proposition 7. *The positive efficiency gain from the existence of the passing activities is achieved only when the passing premium obtained by the passers* ($\int_0^\infty \int_0^{\Delta U} [\Delta U - k] \, \mathrm{d}H(k) \, \mathrm{d}G(c)$) *is greater than the net loss in terms of the reputational externalities, which is the size of negative reputational externalities imposed on the type-B born individuals* ($0.5 \int_0^\infty [U(\pi_l, c) - U(\pi_L^*, c)] \, \mathrm{d}G(c)$) *minus the size of positive reputational externalities imposed on the type-A born individuals* ($0.5 \int_0^\infty [U(\pi_H^*, c) - U(\pi_h, c)] \, \mathrm{d}G(c)$).

The passing premium is directly transformed into the following form, $\int_0^\infty \int_0^{\Delta U} [H(k) - 0.5] \, \mathrm{d}k \, \mathrm{d}G(c)$, in which $\Delta U \equiv \Delta U(\pi_H^*, \pi_L^*; c)$. This implies that the size of the passing premium is largely governed by how much the distribution of the identity manipulation cost is concentrated around zero. That is, the more dense around zero the distribution of identity manipulation cost k is, the greater the societal efficiency gain from the selective out-migration behaviors will be. The positive efficiency gain from the passing activities is more likely to be achieved when identity manipulation is easier to undertake.

It is also notable that the positive societal efficiency gain is achieved when the net loss in terms of the reputational externalities is minimized. Given that $U(\pi, c)$ is a monotonically increasing function of π, roughly speaking, the net loss in terms of the reputational externalities becomes negligible when the fall in reputation of group B ($\pi_l - \pi_L^*$) is not so different from the rise in reputation of group A ($\pi_H^* - \pi_h$) between the two distinct economies.

Furthermore, the negative reputational externalities can even vanish when the disadvantaged group is so severely stigmatized that the believed skill acquisition rate of group B is close to zero in the PSE benchmark economy (i.e., $\pi_l \approx 0$): $0.5 \int_0^\infty [U(\pi_l, c) - U(\pi_L^*, c)] \, \mathrm{d}G(c) \approx 0$ as $\pi_L^* < \pi_l \approx 0$ due to Theorem 1. This implies that the selective out-migration does improve social efficiency when a stigmatized group is severely discriminated against in the labor market equilibrium under the absence of the endogeneities of identity choice.

Corollary 1. *The selective out-migration from a severely stigmatized group with the worst collective reputation* $\pi_l (\approx 0)$ *into any better regarded group with a positive collective reputation* $\pi_h (> 0)$ *always improves social efficiency because the negative reputational externalities produced by the selective migration activities will be negligible, while both the positive reputational externalities and the passing premium are substantial.*

6.2 Developing the Indices of Differentiation ("Partial Passing")

Now consider a stigmatized population for which the pertinent physical traits are not readily disguised or the distinct culture and customs cannot be given up without paying a very high cost (e.g., dark-skinned blacks in the Americas or orthodox Islamic immigrants in Europe). The better-off members of this stigmatized population will not be allowed to pass for a better regarded social group. Instead, they may seek other ways of artful self-presentations to send signals that they are different from the average of the stigmatized mass.

Observers also do not treat all members of the distinct stigmatized population equally. Doing their best under trying circumstances, they may end up partitioning the population in such a way that a person's hard-to-observe but functionally relevant traits (such as whether to make a skill acquisition or not) can be effectively estimated by conditioning on that person's evidently informative though functionally irrelevant traits (such as affectations of speech, dressing up and consumption habits). A "visible" subgroup can be constructed around any cluster of markers shared by some subpopulation. By using a more refined set of indices to guide their discrimination, observers are encouraging the production of those very indices of differentiation by the more talented members of the negatively stereotyped population, who gain most by separating themselves from the mass.

Imagine a specific set of indices that is used for the differentiation. Suppose that employers in the labor market partition the stigmatized population into two subgroups along these indices: subgroup Z' composed of the agents adopting the set of indices and subgroup Z composed of the agents who do not adopt the indices. Assume that the stigmatized population consists of a subpopulation whose natural orientation is to adopt the indices ($k < 0$) and an additional subpopulation whose natural orientation is not to adopt the indices ($k > 0$). An agent with positive k should incur the cost k to be equipped with the indices for the differentiation, while an agent with negative k has no incentive to discard the naturally adopted indices.

The theory developed in this paper is directly applied to this new setting, replacing group A with subgroup Z'. When the believed skill acquisition rate of subgroup Z' is greater than that of subgroup Z, an individual in the stigmatized population can benefit from being perceived as a member of subgroup Z' rather than a member of subgroup Z. The most talented members of the population will disproportionately elect to join the subgroup Z',

adopting the indices, inducing the positive selection into this subgroup and making the human capital cost distributions of the two subgroups diverge endogenously. Denoting the believed skill acquisition rates of the two subgroups by z and z', the stable unequal ESE of (z, z') is (π_L^*, π_H^*), given the existence of multiple PSE (π_l, π_m and π_h), in which $\pi_L^* < \pi_l < \pi_h < \pi_H^*$ holds according to Theorem 1.

The welfare effects of the partial passing behavior can be examined comparing the welfare at the stable ESE (π_L^*, π_H^*) to the welfare at the PSE benchmark economy in which agents do not make a strategic decision on whether to adopt the indices. In this benchmark economy without the strategies of identity manipulation, there should be no difference in terms of the skill acquisition rates between the two subgroup Z and Z' belonging to the stigmatized population: $(z, z') = (\pi_l, \pi_l)$. Refer to Panel B of Figure 5 for this benchmark economy without the partial passing activities being allowed.

Replacing the skill composition in the benchmark economy of the passing scenario (π_l, π_h) with that in the current benchmark economy of the partial passing scenario (π_l, π_l), we obtain the welfare changes of the population aggregates, denoted by $\Delta\ddot{W}_{c,k}$, directly from Proposition 6:

Proposition 8. *Comparing an unequal stable economy with the partial passing activities (π_L^*, π_H^*) with its benchmark economy without such activities (π_l, π_l), "Non-partial passers ($k > \Delta U(\pi_H^*, \pi_L^*; c)$" suffer from the partial passing activities, while the advantaged population ($k < 0$) is benefitted. Whether the welfare of a "partial passer" improves depends on the level of one's identity manipulation cost k. The welfare change of each population aggregate (non-partial passer, partial passer or the advantaged) between the two distinct economies is summarized as*[9]

$$\Delta\ddot{W}_{c,k} = \begin{cases} U(\pi_L^*, c) - U(\pi_l, c), \forall k \in (\Delta U(\pi_H^*, \pi_L^*; c), \infty) & : \textit{negative; "non-partial passers"} \\ U(\pi_H^*, c) - U(\pi_l, c) - k, \forall k \in (0, \Delta U(\pi_H^*, \pi_L^*; c)) & : \textit{positive iff } k < \tilde{k}(c)\textit{; "partial passers"}, \\ U(\pi_H^*, c) - U(\pi_l, c), \forall k \in (-\infty, 0) & : \textit{positive; "the advantaged"} \end{cases}$$

in which there always exist a threshold $\tilde{k}(c)$ $(\equiv U(\pi_H^*, c) - U(\pi_l, c))$ between 0 and $\Delta U(\pi_H^*, \pi_L^*; c)$.

While some partial passers benefit from their own activities, all of the non-partial passers

[9]Note that $\Delta\ddot{W}_{c,k}$ is simply obtained by the replacement of $U(\pi_h, c)$ by $U(\pi_l, c)$ in the equation $\Delta W_{c,k}$ in Proposition 6.

suffer from the behavior of the partial passers. Therefore, such strategies of social identity manipulation through partial passing can be a way to undermine solidarity in the visibly distinct stigmatized population. The worse-off members of the physically marked group (e.g., race-marked population) may accuse the partial passers of some kind of immoral betrayal, which is often referred to as "acting white." Furthermore, the adverse impact on the left-behind may generate the resentment of "acting white." The worse-off members will try to hold such people back by stigmatizing their action to adopt the indices of differentiation.

This is a different explanation of the "acting white" phenomenon from that offered in Austen-Smith and Fryer (2005), which proposes a two-audience model in which the incumbents of the minority population reject their own members who acquire human capital for "acting white" because they think them low social ability types. We suggest that the group reject "partial passers" but not because these people are thought to be socially inept. This group rejects them because it feels betrayed by them and because their departure adversely affects the reputations of those who are left behind (Loury, 2002).

Finally, let us examine whether the partial passing behavior improves social efficiency. The societal efficiency gain from the partial passing activities, which is denoted by $\Delta\ddot{W}_{total}$, is computed by the double integrations of $\Delta\ddot{W}_{c,k}$ in Proposition 8. As demonstrated above in the decomposition of ΔW_{total}, this societal efficiency gain is also composed of two reputational externalities parts (positive and negative) and one passing-related premium part:[10]

$$\begin{aligned}\Delta\ddot{W}_{total} &= \underbrace{0.5\int_0^\infty [U(\pi_H^*, c) - U(\pi_l, c)]\,\mathrm{d}G(c)}_{\text{"positive reputational externality"}} - \underbrace{0.5\int_0^\infty [U(\pi_l, c) - U(\pi_L^*, c)]\,\mathrm{d}G(c)}_{\text{"negative reputational externality"}} \\ &\quad + \underbrace{\int_0^\infty \int_0^{\Delta U} [\Delta U - k]\,\mathrm{d}H(k)\,\mathrm{d}G(c)}_{\text{"partial passing premium"}}.\end{aligned}$$

The change from the PSE benchmark economy (π_l, π_l) to the "partial passing" equilibrium (π_L^*, π_H^*) generates some negative reputational externality for the population aggregate for which the natural orientation is not to adopt the indices for differentiation ($k < 0$), while it may generate some positive reputational externality for the population aggregate for which the natural orientation is to adopt those indices ($k < 0$). While the first and second terms in

[10]Note that $\Delta\ddot{W}_{total}$ is simply obtained by the replacement of $U(\pi_h, c)$ by $U(\pi_l, c)$ in the equation ΔW_{total}.

the above equation indicate the reputational externalities, the third term indicates the partial passing premium for those who choose to adopt the indices of differentiation though incurring some cost for the artful self-presentation.

In general, Proposition 7, which results from the examination of the passing phenomenon, also works for the case of the partial passing phenomenon: The positive efficiency gain is achieved only when the partial passing premium obtained by the partial passers is greater than the net loss in terms of the reputational externalities. One key difference is that the chance for the positive efficiency gain is much higher for this case because the positive reputational externality has increased significantly by as much as $0.5\int_0^\infty [U(\pi_h, c) - U(\pi_l, c)]\,\mathrm{d}G(c)$ compared to the positive externalities from the passing activities.

As noted in Corollary 1, the negative reputational externality generated from the partial passing activities will be negligible when the stigmatized population is severely discriminated in the labor market equilibrium under the absence of artful self-presentation ($\pi_l \approx 0$): $0.5\int_0^\infty [U(\pi_l, c) - U(\pi_L^*, c)]\,\mathrm{d}G(c) \approx 0$ as $\pi_L^* < \pi_l \approx 0$. Therefore, a positive efficiency gain from the partial passing activities is automatically ensured in this case. The development of the indices of differentiation by the better-off members of the severely stigmatized group improves the societal efficiency without hurting the welfare of the left-behind members.

7 Conclusion

Our theoretical model in this paper is based on a stereotyping-cum-signaling framework suggested by Arrow (1973) and Coate and Loury (1993), in which multiple self-confirming beliefs by employers about different social identity groups explain the between-group inequality in terms of the skill acquisition activities. Unlike the previous works and their subsequent developments (e.g., Moro and Norman, 2004; Chaudhuri and Sethi, 2008), we handle the dynamics between the collective reputation and the identity choice problem. By relaxing the immutability assumption, the model explores the implications of the fact that the distribution of abilities within distinct identity groups becomes endogenous when individuals choose how they will be identified by external observers.

The main result is that the group inequality that derives from the ascriptive stereotyping of endogenously constructed social groups is at least as great as the inequality that can emerge from the phenotypic stereotyping of exogenously given groups. In the standard statistical

discrimination literature, the inequality between groups is due to the positive complementarities between a group's reputation and its members' investment activities. In the current framework that allows observed agents to exert control over their perceived identities, the inequality is not only due to the positive complementarities but also due to the positive selection along the ability parameter. The low human capital cost types are disproportionately drawn to the group with a better collective reputation, causing a skill disparity between groups to endogenously diverge.

The model developed in this paper is a step toward a better understating of identity choice behaviors. We have applied the theory to the passing and 'partial passing' phenomena. We find that these identity manipulation activities can improve the social efficiency either when the (partial) passing premium is maximized or when the loss in terms of the reputational externalities is minimized. The former is maximized when identity manipulation is easier to undertake, and the latter is minimized when the fall in reputation of the disadvantaged group is not much greater than the rise in reputation of the advantaged group. Identifying who benefits and who suffers from the phenomena, we provide the rationale behind conflicts within a stereotyped population, i.e., the 'acting white' accusation. The theory also has the potential to illuminate many different social phenomena related to the choice of the perceived identities, which include the use of racial profiling in law enforcement, the coming out decision by LGBT people and the marketing problem of effectively 'branding' a new consumer product.

8 Appendix: Proofs

8.1 Proof of Lemma 2

Consider a very small $\delta > 0$ such that $a = b + \delta$. Define $\Delta(\delta; b)$ as $\Delta(\delta; b) \equiv R(b+\delta) - R(b)$, which is a function of δ given b: $\Delta' \left(\equiv \frac{\partial \Delta(\delta;b)}{\partial \delta}\right) = R'(b+\delta)$. We have $H'(k) \approx H'(0)$ for small enough k. Using Lemma 1 and Panels A and B of Figure 2, we can develop $\sigma^A(\delta; b)$ and $\Sigma^A(\delta; b)$, which are functions of δ given b, and consequently $\sigma^{A'}(\delta; b)$ and $\Sigma^{A'}(\delta; b)$. The fraction of agents who elect to be A-type and decide to be skilled is

$$\begin{aligned}
\sigma^A(\delta; b) &\approx G(R(b)+\Delta) \cdot [.5 + H'(0)(V_1(b+\delta) - V_1(b))] - .5H'(0)g(R(b))\Delta^2, && (21)\\
\sigma^{A'}(\delta; b) &\approx g(R(b)+\Delta)R'(b+\delta)[.5 + H'(0)(V_0(b+\delta) - V_0(b))] \\
&\quad + G(R(b)+\Delta)H'(0)V_1'(b+\delta) - H'(0)g(R(b))\Delta R'(b+\delta), && (22)
\end{aligned}$$

in which the last terms that represent a triangle area, $-.5H'(0)R'(b)\Delta^2$ and $-H'(0)g(R(b))\Delta R'(b+\delta)$, are added when $R'(b) > 0$ (as in Panel A), and dropped when $R'(b) < 0$ (as in Panel B). The fraction of agents who elect to be A-type is

$$\begin{aligned}
\Sigma^A(\delta; b) &\approx .5 + H'(0)(V_0(b+\delta) - V_0(b)) + G(R(b)+\Delta)H'(0)\Delta - .5H'(0)g(R(b))\Delta^2, && (23)\\
\Sigma^{A'}(\delta; b) &\approx H'(0)V_0'(b+\delta) + G(R(b)+\Delta)H'(0)R'(b+\delta) + g(R(b)+\Delta)R'(b+\delta)H'(0)\Delta \\
&\quad - H'(0)g(R(b))\Delta R'(b+\delta). && (24)
\end{aligned}$$

From the approximations with small δ given b, we can achieve the following results when $\delta = 0$:

$$\begin{cases}
\sigma^A(0; b) & \approx .5G(R(b)) \\
\sigma^{A'}(0; b) & \approx .5g(R(b))R'(b) + G(R(b))H'(0)V_1'(b) \\
\Sigma^A(0; b) & \approx .5 \\
\Sigma^{A'}(0; b) & \approx H'(0)V_0'(b) + G(R(b))H'(0)R'(b)
\end{cases} \tag{25}$$

The slope of the $\phi(a;b)$ curve given a=b can be expressed as follows, using $\sigma^A(\delta;b)$ and $\Sigma^A(\delta;b)$,

$$\begin{aligned}
\frac{\partial\phi(a;b)}{\partial a}\Big|_{a=b} &= \lim_{\delta\to 0}\frac{\phi(b+\delta;b)-\phi(b;b)}{\delta} \\
&= \lim_{\delta\to 0}\frac{\sigma^A(\delta;b)/\Sigma^A(\delta;b)-\sigma^A(0;b)/\Sigma^A(0;b)}{\delta} \\
&= \lim_{\delta\to 0}\left[\frac{[\sigma^A(\delta;b)-\sigma^A(0;b)]\Sigma^A(0;b)}{\delta}-\frac{[\Sigma^A(\delta;b)-\Sigma^A(0;b)]\sigma^A(0;b)}{\delta}\right] \\
&\quad \cdot\frac{1}{\Sigma^A(\delta;b)\Sigma^A(0;b)} \\
&\approx \frac{\sigma^{A'}(0;b)\Sigma^A(0;b)-\sigma^A(0;b)\Sigma^{A'}(0;b)}{\Sigma^A(0;b)^2} \qquad (26)
\end{aligned}$$

Applying the above results from approximations and noting $R'(b)\equiv V_1'(b)-V_0'(b)$, we achieve that:

$$\frac{\partial\phi(a;b)}{\partial a}\Big|_{a=b}\approx g(R(b))R'(b)+2H'(0)R'(b)G(R(b))\cdot[1-G(R(b))]. \qquad (27)$$

QED.

8.2 Proof of Lemma 4

We can find a correspondence value x' nearby $\hat{x}$ such that $x'=\phi(x';\hat{x}+\Delta)$, which means $x'\in\Gamma(\hat{x}+\Delta)$, as displayed in Appendix Figure 1. Given the slope of $\phi(x;y)$ at $(\hat{x}+\Delta,\hat{x}+\Delta)$ denoted by $\frac{\partial\phi(x;y)}{\partial x}\big|_{x=y=\hat{x}+\Delta}$ and the slope of $\phi(x;x)$ at $(\hat{x},\hat{x})$, $g(R(\hat{x}))R'(\hat{x})$, the correspondence value x' approximately satisfies the following equation:

$$x'-[\hat{x}+g(R(\hat{x}))R'(\hat{x})\Delta]\approx\frac{\partial\phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta}\cdot[x'-(\hat{x}+\Delta)]\,. \qquad (28)$$

The slope of the correspondence curve at a trivial ESE $(\hat{x},\hat{x})$, denoted by $\Gamma'(\hat{\pi})$, is approximately equal to $\lim_{\Delta\to 0}\frac{x'-\hat{x}}{\Delta}$:

$$\Gamma'(\hat{x})\approx\left[g(R(\hat{x}))R'(\hat{x}))-\lim_{\Delta\to 0}\frac{\partial\phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta}\right]\Big/\left[1-\lim_{\Delta\to 0}\frac{\partial\phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta}\right]. \qquad (29)$$

From Lemma 2 and $G(R(\hat{x}))=\hat{x}$, we have

$$\lim_{\Delta\to 0}\frac{\partial\phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta}=g(R(\hat{x}))R'(\hat{x})+2H'(0)R'(\hat{x})\hat{x}(1-\hat{x}). \qquad (30)$$

Then, we have the given result for $\Gamma'(\hat{x})$. Q.E.D.

8.3 Proof of Proposition 3

Given multiple PSE (π_l, π_m, π_h) and the condition of $\pi_h < \bar{\pi}$, $\Gamma(b)$ passes through the points (π_h, π_h) and a-intercept $(b, a) = (0, \Gamma(0)^h)$, in which $\pi_h < \Gamma(0)^h < 1$. $\Gamma(a)$ passes through (π_l, π_l) and b-intercept $(a, b) = (1, \Gamma(1)^h)$, in which $\pi_h < \Gamma(1)^h < 1$. Thus, there should be at least one ESE (b^*, a^*) that satisfies $a^* > b^*$. In a symmetric way, we can prove the existence of at least one ESE that satisfies $b^* > a^*$. Q.E.D.

8.4 Proof of Theorem 1

Given multiple PSE, we can find that $\pi_h < \Gamma(\pi_l)^h < \tilde{b}$, in which $\phi(\tilde{b}; \pi_l) = \pi_l$ as indicated in Figure 3. We can draw the shape of $\phi(a; \Gamma(\pi_l)^h)$, which passes through the 45 degree line at some $a < \pi_l$. This implies that $\Gamma(\Gamma(\pi_l)^h)^l < \pi_l$. Because $\Gamma(b)^h$ decreases over $b \in (0, \bar{\pi})$, there must be an intercept of $\Gamma(b)$ and $\Gamma(a)$, (π_L^*, π_H^*), which satisfies both $\pi_L^* < \pi_l$ and $\pi_H^* > \pi_h$. Out of the symmetry, there must be an additional ESE (π_H^*, π_L^*). Q.E.D.

References

Anderson, Elijah. *Streetwise: Race, Class, and Change in an Urban Community.* Chicago: University of Chicago Press, 1990.

Arrow, Kenneth. "Political and Economic Evaluation of Social Effects and Externalities," in Michael Intriligator, ed., *Frontiers of Quantitative Economics.* Amsterdam: North-Holland, 1971, pp. 841–77.

———. "The Theory of Discrimination," in Ashenfelter and Rees, ed., *Discrimination in Labor Markets.* 1973.

Charles, Kerwin K., Erik Hurst and Nikolai Roussanov. "Conspicuous Consumption and Race." *The Quarterly Journal of Economics*, 2009, 124(2), pp. 425–67.

Chaudhuri, Shubham and Rajiv Sethi. "Statistical Discrimination with Peer Effects: Can Integration Eliminate Negative Stereotypes?" *Review of Economic Studies*, 2008, 75(2), pp. 579–96.

Coate, Stephen and Glenn Loury. "Will Affirmative-Action Policies Eliminate Negative Stereotype?" *American Economic Review*, 1993, 83(4), pp. 1220–40.

Fang, Haming. "Social Culture and Economic Performance." *American Economic Review*, 2001, 91(4), pp. 924–37.

Fiske, Susan. "Stereotyping, prejudice, and discrimination," in D. T. Gilbert, S. T. Fiske and G. Lindzey, ed., *Handbook of Social Psychology.* New York: McGraw-Hill, 1998, 2, pp. 357-411.

Fukuoka, Yasunori and Yukiko Tsujiyama. "Mintohren: Young Koreans against Ethnic Discrimination in Japan." *The Bulletin of Chiba College of Health Science* 1998, 10 (2), pp. 147–62.

Goffman, Erving. *The Presentation of Self in Everyday Life.* New York: Doubleday, 1959.

———. *Stigma: Notes on the Management of Spoiled Identity.* Prentice-Hall, 1963.

Greenwald, Anthony and Mahzarin Banaji. "Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes." *Psychological Review*, 1995, 102(1), pp. 4–27.

Grogger, Jeffrey. "Speech Patterns and Racial Wage Inequality." *The Journal of Human Resources*, 2011, 46(1), pp. 1–25.

Kim, Young-Chul and Glenn Loury. "Stereotypes and Inequality: A 'Signaling' Theory of Identity Choice." *The KDI Journal of Economic Policy*, 2012, 34(2), pp. 1–15.

Loury, Glenn. *The Anatomy of Racial Inequality.* Harvard University Press, 2002.

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

Sampson, Robert and Stephen Raudenbush. "Seeing Disorder: Neighborhood Stigma and the Social Construction of "Broken Windows"." *Social Psychology Quarterly*, 2004, 67(4), pp. 319–42.

Srinivas, Mysore. *Religion and Society among the Coorgs of South India.* Oxford University Press, 1952.

Steele, Claude and Joshua Aronson. "Stereotype Threat and the Intellectual Test Performance of African-Americans." *Journal of Personality and Social Psychology*, 1995, 69(5), pp. 797–811.

Sweet, Frank. "The Rate of Black-White "Passing"." unpublished manuscript, 2004.

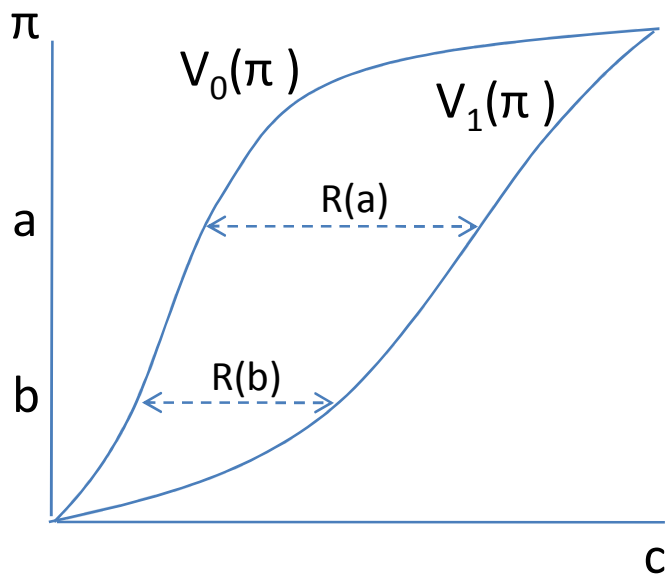
Tajfel, Henri. "Social Identity and Intergroup Behavior." *Social Science Information*, 1974, 13(2), pp. 65-93.

Telles, Edward. "Racial Classification." *Race in Another America: the Significance of Skin Color in Brazil.* Princeton University Press, 2004, pp. 81-84

Tirole, Jean. "A Theory of Collective Reputation." *Review of Economic Studies*, 1996, 63(1), pp. 1–22.

Figure 1. Phenotypic Stereotyping Equilibria

Panel A. Expected Rewards Given π



Panel B. Multiplicity of Equilibria

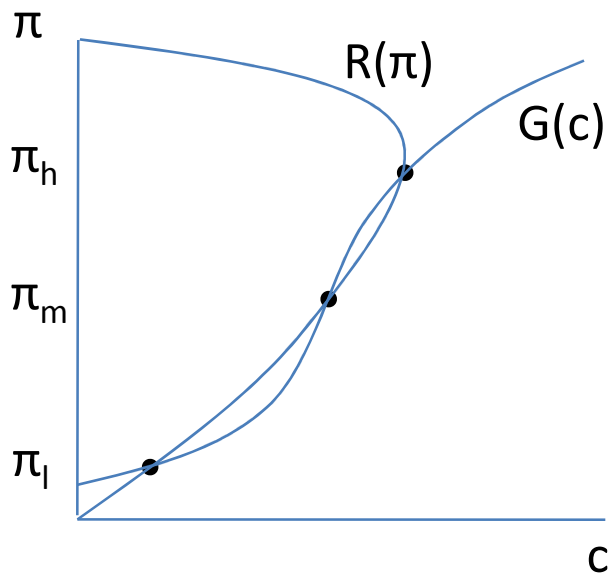
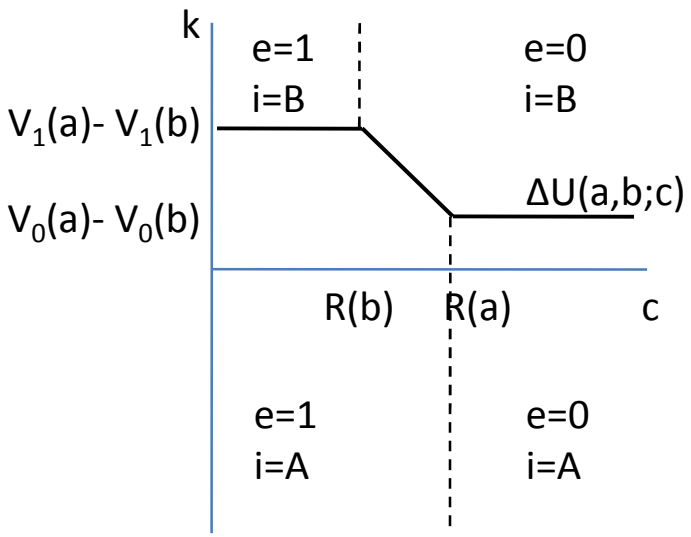
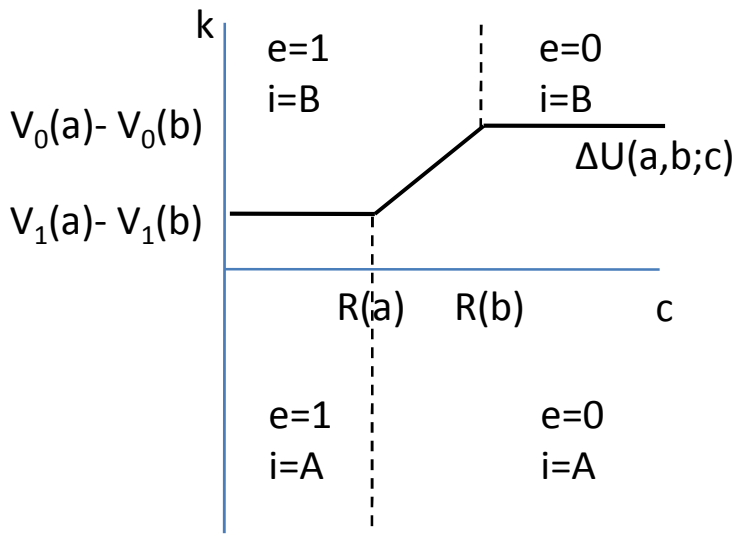


Figure 2. Human Capital Investment and Identity Choice Behavior

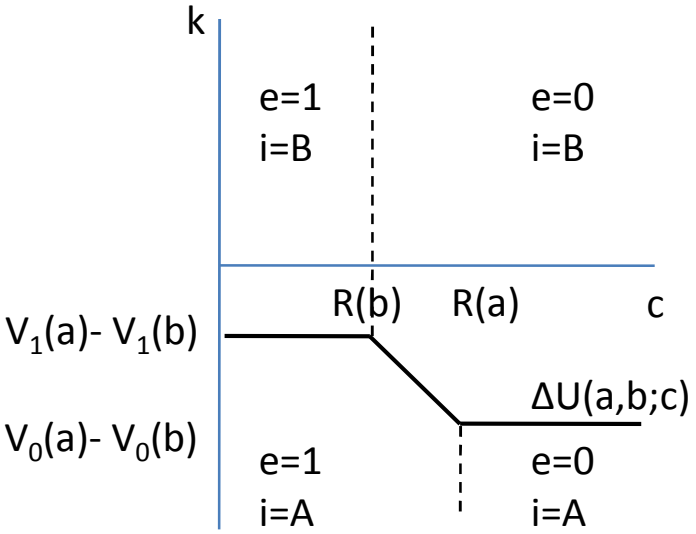
Panel A. Case with a>b and R(a)>R(b)



Panel B. Case with a>b but R(a)<R(b)



Panel C. Case with a<b but R(a)>R(b)



Panel D. Case with a<b and R(a)<R(b)

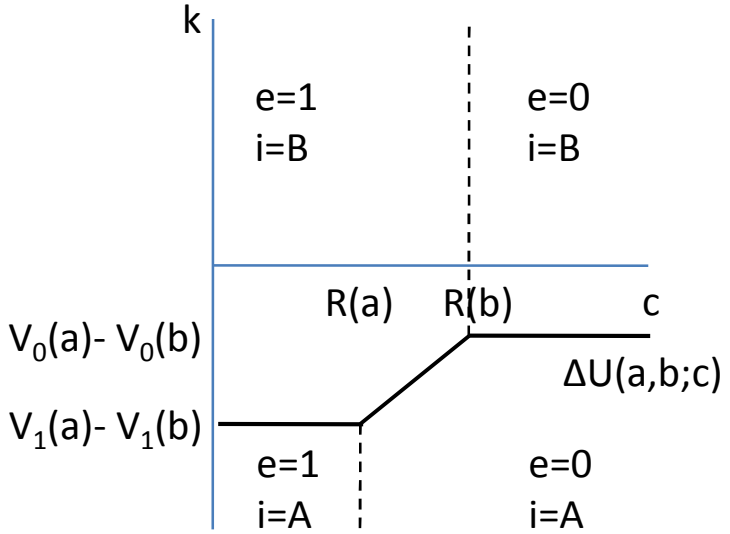


Figure 3. Human Capital Investment Rate $\phi(a; b)$

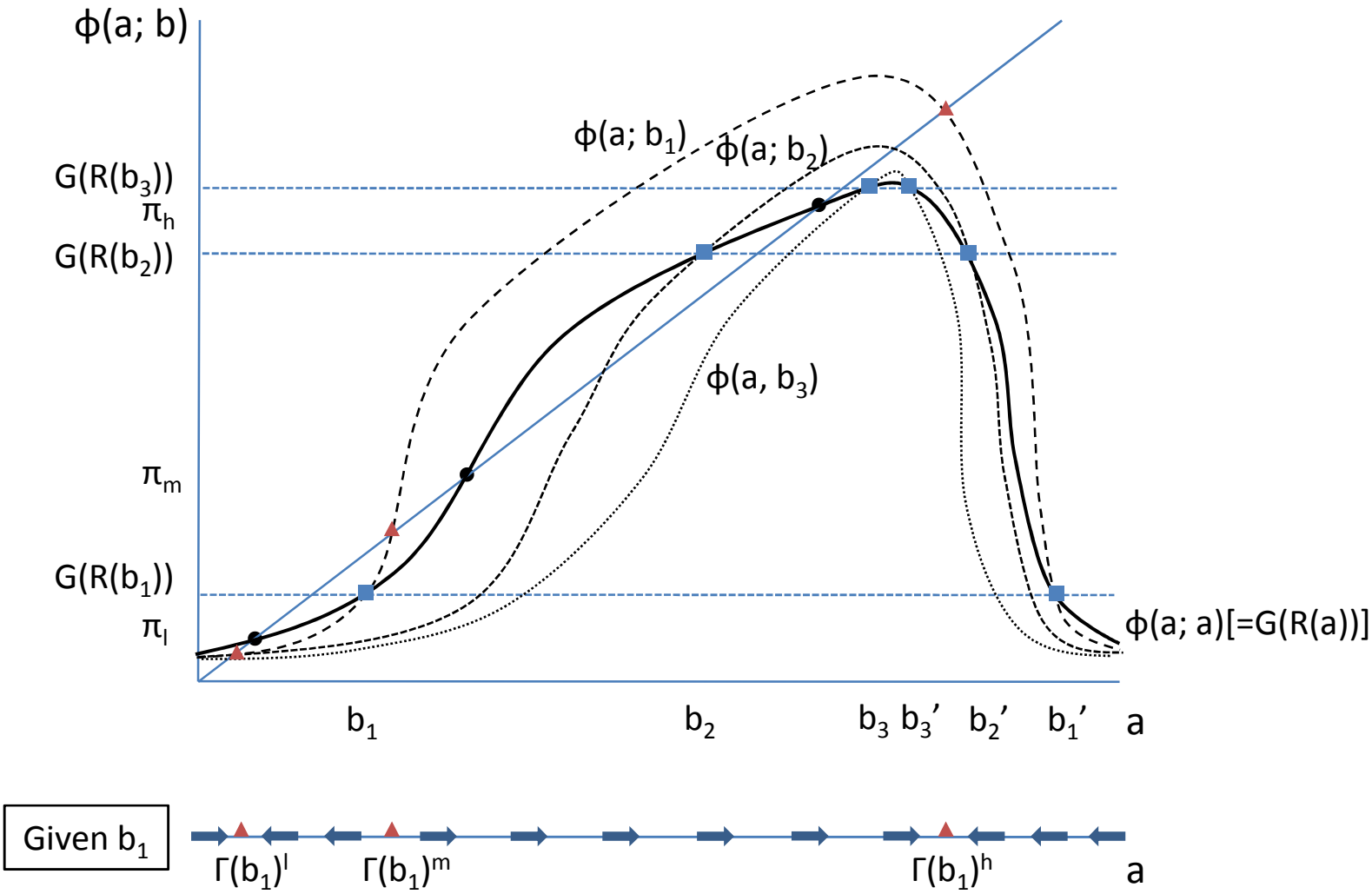
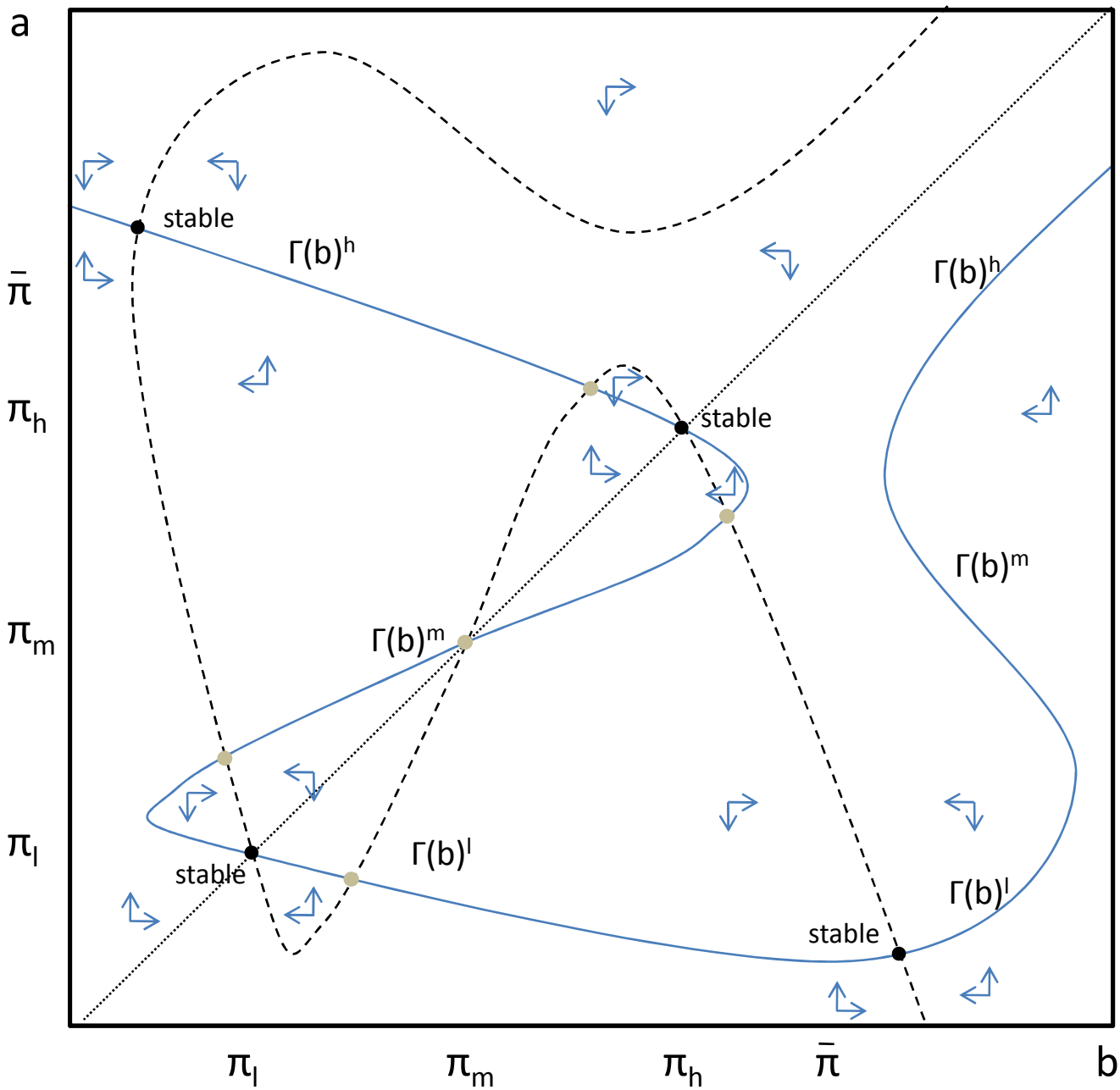
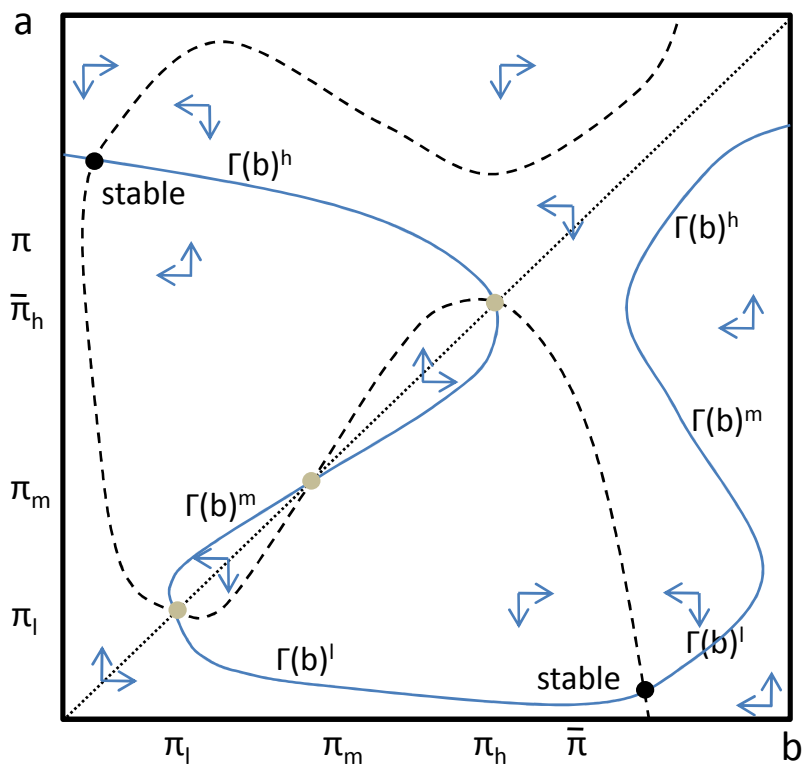


Figure 4. ESE given Multiple PSE (π_l, π_m, π_h)

Panel A. Given both $-1<\Gamma'(\pi_h)<0$ and $-1<\Gamma'(\pi_l)<0$



Panel B. Given both $\Gamma'(\pi_h)<-1$ and $\Gamma'(\pi_l)<-1$



Panel C. Given both $\Gamma'(\pi_h)>1$ and $\Gamma'(\pi_l)>1$

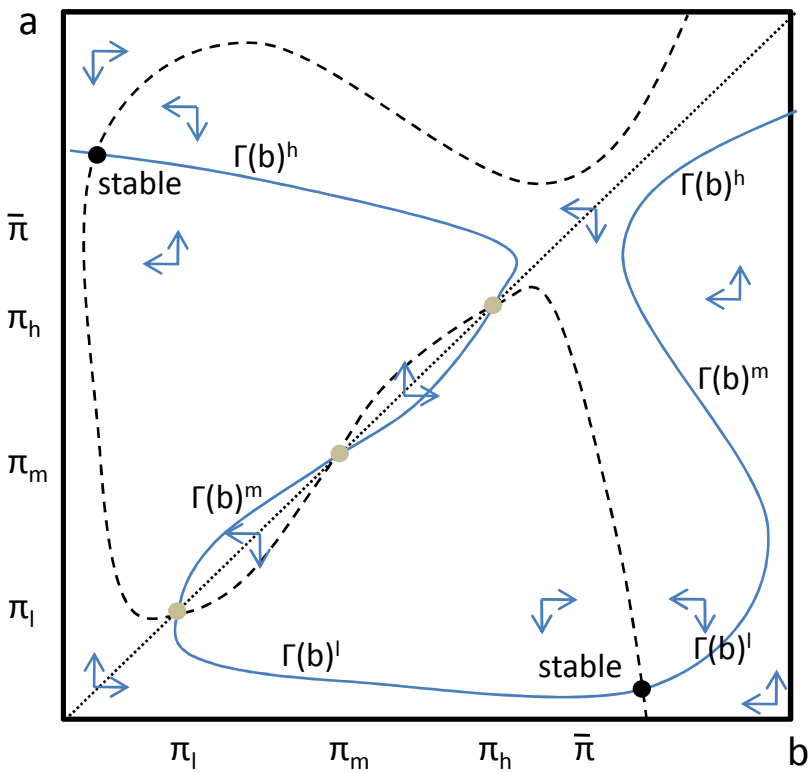
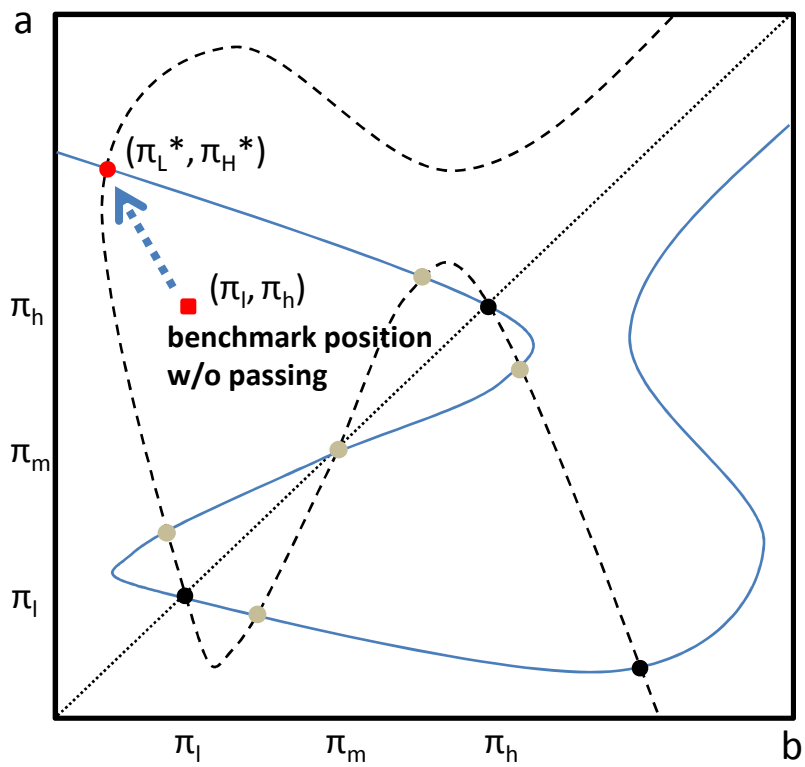


Figure 5. Passing and Partial Passing Behaviors

Panel A. Passing (Group B to Group A)



Panel B. Partial Passing (Subgroup Z to Subgroup Z′)

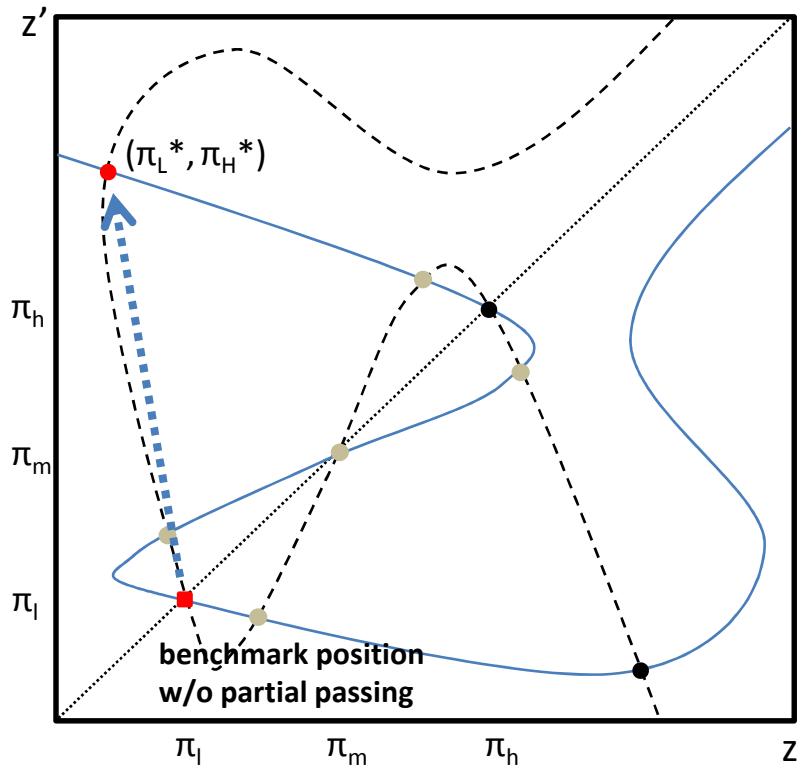
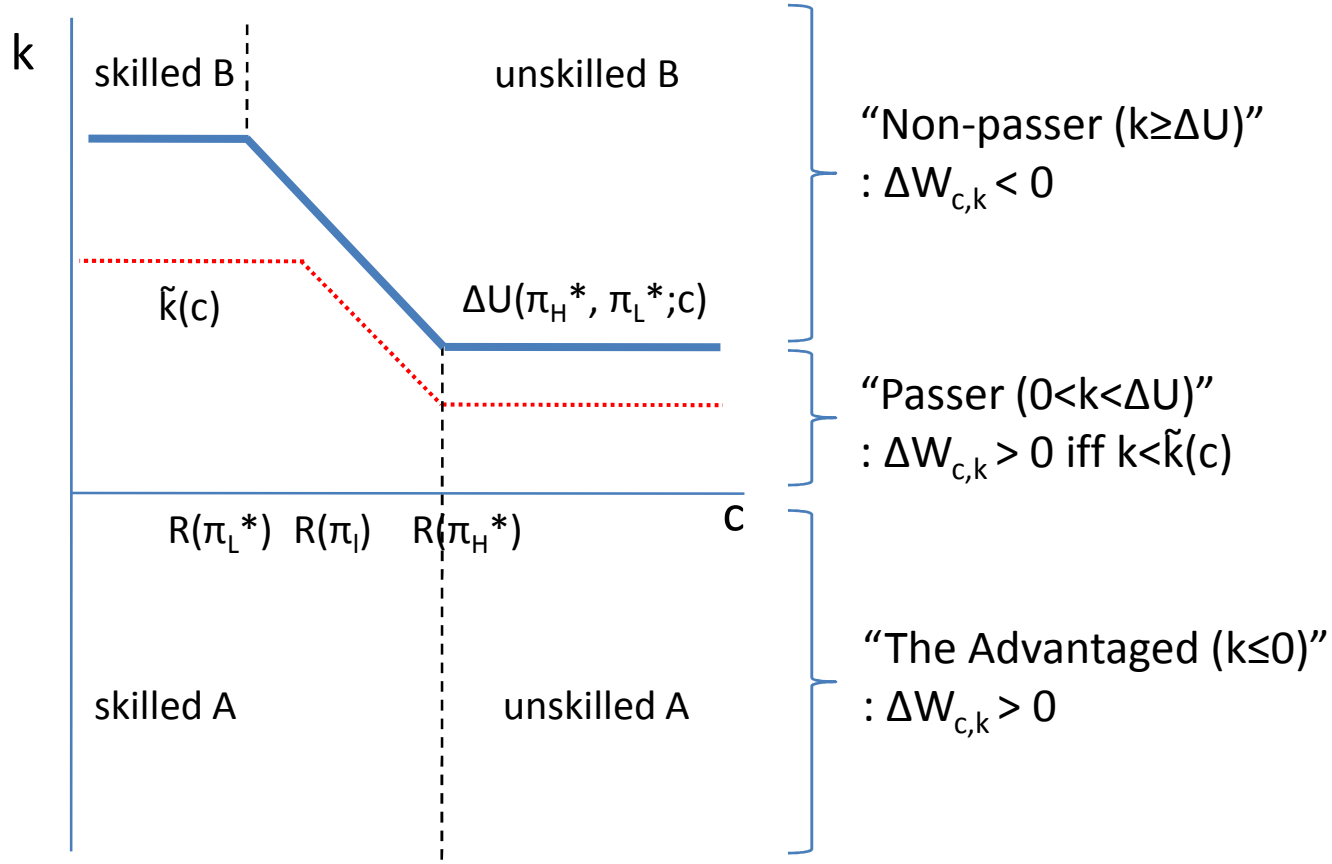
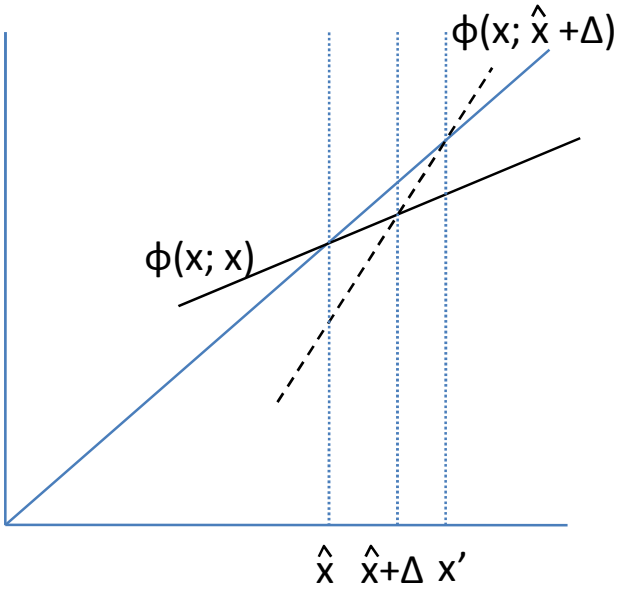
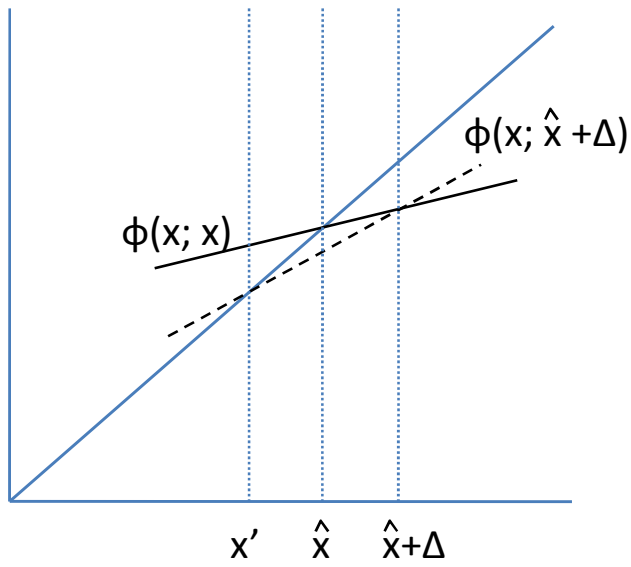


Figure 6. Welfare Effects of Passing



Appendix Figure 1. Slope of Correspondence at Trivial ESE

Panel A. Example for $\Gamma'(\hat{x})>0$



Panel B. Example for $\Gamma'(\hat{x})<0$