Stereotypes and Identity Choice

Young-Chul Kim[†] Glenn C. Loury[‡]

April 15, 2016

Abstract

We extend a model of 'stereotyping' by allowing agents to exert control over their perceived identities. The logic of individuals' identity choices induces 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 groups is at least as great as the inequality that can emerge when perceived identity is not malleable. Among the human behaviors illuminated by this theory are: (1) the selective outmigration from a stigmatized group and (2) the production of the indices of differentiation by better-off members of the negatively stereotyped group.

KEYWORDS: Stereotypes, Identity Choice, Group Inequality, Passing.
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, 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 72th Annual Conference of the Japan Institute of Public Finance (2015). We are indebted to Rajiv Sethi, Matthew Jackson, Alan Kirman and Samuel Bowles for their insightful comments. All remaining errors are ours.

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

[‡]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 on the basis of their common possession of visible marks or other observable characteristics, 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. The 'collective reputations' are this sort of rational formation by external observers of beliefs about the unobserved traits of varied population aggregates. This 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; Chaudhuri and Sethi, 2008), 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. 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 pioneered 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 an economically relevant dimension such as ability 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 identity group may vary 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 clearly 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 human 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, which is termed as 'partial passing' in this paper. The examples relevant for these phenomena are introduced extensively in the next section.

In our theoretical framework, we define two distinct equilibria: PSE and ESE. A standard statistical discrimination framework (e.g., Coate and Loury, 1993) entails no selection into or out of the groups. We call the self-confirming belief equilibrium with exogenous social identities a 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. We call such a group-disparate equilibrium with endogenous identities an Endogenous Stereotyping Equilibrium (ESE).[1]

Comparing PSE and ESE, 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

[1] In search for the equilibrium, 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.

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. We prove the existence and the stability of such unequal ESE, given the presence of multiple PSE. In addition, those unequal ESE are the only stable equilibria when identity manipulation is sufficiently 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). Therefore, these 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 an alternative explanation of the 'acting white' phenomenon to those offered by other scholars (e.g., Austen-Smith and Fryer, 2005).

Through the decomposition of the societal efficiency gain into reputational externalities and passing (partial passing) premium, however, we show that these identity manipulation activities can increase the total welfare of the society under some limited conditions. Furthermore, we demonstrate that when a stereotyped group is severely discriminated against, the activities can improve the societal efficiency even without hurting the welfare of the left-behind.

Various concepts of "identity" have been developed recently in the growing literature on the economics of identity: e.g., a person's sense of self that affects preferences (Akerlof and Kranton, 2000), moral identity as beliefs about one's deep "value" (Benabou and Tirole, 2011), group identity as a shared convention (Fang and Loury, 2005), and differences in norms tied to social identities (Benjamin, Choi and Strickland, 2010). Our approach to the concept of "identity" differs from theirs, in the sense that it represents a social group's collective reputation, which in turn affects labor market outcomes. Some of the men-

tioned works also discuss the identity choice problem, but from a viewpoint very different from ours.

The remainder of this paper is structured as follows. Section 2 introduces various real life examples that are relevant for passing and 'partial passing' phenomena. Section 3 describes the basic structure of the signaling model, in which agents decide on the perceived identity as well as the skill acquisition. Section 4 defines both PSE and ESE. Section 5 studies the properties of the identity choice behaviors, and Section 6 examines the existence and the stability of ESE. Section 7 follows with a discussion of the welfare properties of the equilibria. Section 8 presents the study's conclusion.

2 Examples of Identity Choice Behaviors

Young members in a stereotyped 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.) These stereotyped social 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 and Israel, along ethnic lines in Singapore and the Balkan states, with caste-like social division in the Indian sub-continent and the treatment of Gypsies and immigrants in Europe. The selective out-migration occurs as more talented members in the disadvantaged groups cross the color/religious/ethnic/caste lines disproportionately.

A manifest 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.[2] To escape the negative stereotypes and prejudices against the

[2]Every year, approximately 10,000 Koreans, of approximately 600,000 Korean descendants holding Korean nationality, choose to be naturalized as 'official' Japanese mostly when seeking formal employment or marriage.

Zainichi, many of the naturalized Koreans conceal their ethnicity, giving up their names and pretending that they have no knowledge about Korean culture and language (Fukuoka et al., 1998).

Other than the Zainichi, who share a similar appearance with the Japanese, passing is harder for blacks and other minorities in the United States due to their physical makeup. However, the light-skinned minorities with mixed ancestry have been crossing the boundaries of color and racial identity.[3] In old Hollywood, for instance, talented movie stars were expected to downplay their ethnic origins when they were not solely of European extraction.[4]

Unlike the United States, which had defined concepts of race due to the 'one drop rule,' 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 countries, some of which might be classified as 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. The journalists report that the fascination with becoming "white" has increased over the last decades with the prevailing "whitening" practices (e.g., the use of skin bleaching cosmetics and the treatments to straighten hair) among the mixed-race youngsters.

In other situations, discriminated groups may modify their accents, word choices, manner of dress and even custom in an attempt to appear to be members of a privileged group.[5] This type of passing in the context of caste is called

[3]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).

[4]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). Besides, it was not uncommon for stars of even European extraction to downplay their roots by adopting American sounding names.

[5]For example, My Fair Lady, a musical based upon George Bernard Shaw's *Pygmalion*,

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 (Srinivas, 1952).[6]

Passing into the better-regarded group is not always possible for every stigmatized group. 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, whereby people with any known African ancestry were automatically classified as Black, 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).

There is systematic empirical evidence regarding the styles of self-presentation for social identity manipulation. For instance, 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. Grogger (2011) 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

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.

[6] 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.

percent less than whites with similar observable skills, while indistinctly identified blacks earn essentially the same as comparable whites. Such speech-related wage premia may provide incentives for the talented blacks to adopt standard American English rather than African American English. Then, in the labor market, speech patterns can signal the worker's underlying abilities.[7]

The theoretical model developed below explains the rationale behind these identity choice behaviors and explores implications of such fact that the distribution of abilities within distinct identity groups becomes endogenous.

3 Framework of the Model

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 are matched with an employer.

A worker's skill acquisition decision is denoted by $e \in \{0, 1\}$. The cost of obtaining a skill varies among the workers: $c \in [0, \infty]$. Workers 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 c. The cost distribution $G(c)$ satisfies $G(0) > 0$, implying the existence of a fraction of highly capable workers whose skill acquisition cost is sufficiently low. 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 $\hat{c}$ (e.g., normal distribution).[8] An agent with cost c invests in skills if and only if the anticipated return from

[7]Charles et al. (2009) made a careful examination of the Consumer Expenditure Survey (CES) by the U.S. Department of Labor. Grogger (2011) used audio data from interviews administered to the National Longitudinal Survey of Youth (NLSY) respondents.

[8]Most ability-related test scores reveal single-peaked distributions of intelligence. For instance, SAT scores are approximately normally distributed over the tested population. The widely-used intelligence quotient (IQ) scores are also distributed normally about 100, with a standard deviation of 15. If a person's intelligence is affected by a large number of independent causes, each of which has a small effect, intelligence can be argued to be distributed normally across the population (Hunt, 2011).

doing so exceeds this cost for the skill acquisition.

The workers are also allowed to choose, prior to being matched with an employer, how they are categorized and perceived in the labor market. 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, so called identity "switching" cost, is $k \in R$. This can be positive or negative.[9] 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. Therefore, the variable k successfully reflects a cost to adopting a different identity than one's own natural one. The cumulative distribution function (CDF) of the cost is denoted by $H(k)$. For the sake of simplicity, we assume the symmetry of the distribution: $H(k) = 1 - H(-k)$.[10] 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, implying that a person's identity orientation cannot be used to predict his or her economic ability.

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 e 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

[9] If k is negative, the relative "benefit" of being perceived as A rather than B is $-k$.

[10] 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.

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 $t \in (0,1)$.[11] 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: $\rho(\pi,t)(\equiv Pr[e=1|\pi,t]) = \frac{\pi f_1(t)}{f(\pi,t)}$. We assume 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 competitive wage denoted by W will be the workers' expected productivity: $W \equiv w \cdot \rho(\pi,t)$. Then, the anticipated wage 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

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

Given this framework, we can readily express the expected payoff from acquiring a skill ($e=1$) and that without acquiring a skill ($e=0$) as follows:

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

[11] 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 $F_1(t) - F_0(t) = -\int_{\bar{t}}^1 f_1(x)(1-\frac{f_0(x)}{f_1(x)})\,\mathrm{d}x < 0$.

in which both derivatives $V_0'(\pi)$ and $V_1'(\pi)$ are always positive, indicating that they are increasing functions of the believed skill acquisition rate π, as depicted in Panel A of Figure 1.[12] We can also derive that $\lim_{\pi\to 0} V_0'(\pi) = w$ and $\lim_{\pi\to 1} V_1'(\pi) = w$.

Workers' expected economic return from being skilled, which is denoted by $R(\pi)$, is equivalent to the difference between the expected payoff from acquiring a skill and that without acquiring a skill: $R(\pi) \equiv V_1(\pi) - V_0(\pi)$. Given $\pi = 0$, both the expected payoff from acquiring a skill and that without acquiring a skill are zero, implying that the expected economic return from being skilled is zero: $V_1(0) = V_0(0) = 0$ and $R(0) = 0$. Using a similar logic, given $\pi = 1$, we have $V_1(1) = V_0(1) = w$ and $R(1) = 0$.

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

$$R(\pi) = w\pi \int_0^1 \frac{(f_1(t) - f_0(t))f_1(t)}{f(\pi, t)}\,\mathrm{d}t. \tag{3}$$

The first and second derivatives of the return function can be directly seen as:

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

$$R''(\pi) = -2w \int_0^1 \frac{(f_1(t) - f_0(t))^2 f_1(t)f_0(t)}{f(\pi, t)^3}\,\mathrm{d}t. \tag{5}$$

Using MLRP property, we can derive that $\lim_{\pi\to 0} R'(\pi) > 0$ and $\lim_{\pi\to 1} R'(\pi) < 0$.[13] Because 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 $R'(\bar{\pi}) = 0$.

[12] Note that the first derivatives are derived as $V_0'(\pi) = \int_0^1 w f_1(t) f_0(t)^2 f(\pi, t)^{-2}\,\mathrm{d}t$ and $V_1'(\pi) = \int_0^1 w f_1(t)^2 f_0(t) f(\pi, t)^{-2}\,\mathrm{d}t$.

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

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

$$U(\pi, c) = \max\{V_1(\pi) - c,\ V_0(\pi)\}, \tag{6}$$

in which the function $U(\pi, c)$ is increasing in π for both $V_1(\pi)$ and $V_0(\pi)$ are increasing in π. The function is non-increasing in c given any fixed level of π.

4 Phenotypic vs Endogenous Stereotyping Equilibria

In this section, we define both the Phenotypic and Endogenous Stereotyping Equilibria given the above theoretical framework.

4.1 Phenotypic Stereotyping Equilibria

Imagine 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. 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).

4.2 Endogenous Stereotyping Equilibria

Now 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 net reward 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 type-B 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 also that $\Delta U(a,b;c) = -\Delta U(b,a;c)$ and $\Delta U(a,b;c) = 0$ when $a = b$.

An agent with the endowed cost set (c,k) elects to be an A-type worker if and only if $k \leq \Delta U(a,b;c)$. Otherwise, he elects to be a B-type worker. 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 $G(c)$,

$$\Sigma^A \equiv \int_0^\infty H(\Delta U(a,b;c))\,\mathrm{d}G(c). \tag{7}$$

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 (i.e., $c \leq R(a)$) will elect to be skilled. Then, the fraction of workers who elect to be A-type and become skilled is given by

$$\sigma^A \equiv \int_0^{R(a)} H(\Delta U(a,b;c))\,\mathrm{d}G(c). \tag{8}$$

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

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

Consequently, the fraction of workers who elect to be B-type and become skilled is given by

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

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:

$$\phi(x;y) \equiv \frac{\int_0^{R(x)} H(\Delta U(x,y;c))\,\mathrm{d}G(c)}{\int_0^{\infty} H(\Delta U(x,y;c))\,\mathrm{d}G(c)}, \tag{11}$$

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}.

4.3 Correspondence and the Set of Equilibria

In order to analyze the equilibria effectively, we introduce a correspondence $\Gamma(y)$: $\Gamma(y) = \{x \,|\, x = \phi(x;y)\}$. 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 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 $y = \phi(y;x)$. This also implies that every PSE corresponds to trivial ESE where differences in affect are uninformative: $(\hat{x},\hat{x}) \in \Omega_{KL}$ if $\hat{x} \in \Psi_{CL}$.

Before we start to search for PSE/ESE in the given framework, readers may

review Appendix A first to grasp the key mechanism and the main intuitions of the model, in which those equilibria are determined in a setup with the simplest possible cost structures: agents are composed of only three types of human capital cost (c_l, c_m, c_h) and only four types of identity manipulation cost $(K_l, K_h, -K_l, -K_h)$. The set of ESE in such setup is depicted in Panel D of Appendix Figure 1: given two PSE, Π_l and Π_h, there exist two non-trivial ESE, (Π'_l, Π'_h) and (Π'_h, Π'_l), which satisfy $\Pi'_l < \Pi_l < \Pi_h < \Pi'_h$, implying that the inequality between endogenously constructed groups can be greater than the inequality that can emerge between exogenous groups.

5 Properties of Identity Choice Behaviors

In this section, we examine the key properties of the identity choice behaviors in the given original framework. Acknowledge that the expected net reward 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 (6),

$$\Delta U(a, b; c) = \max\{R(a), c\} - \max\{R(b), c\} + V_0(a) - V_0(b). \tag{12}$$

This expression helps us to achieve the following lemma concerning the varying values of $\Delta U(a, b; c)$:

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*

$$\Delta U(a, b; c) = \begin{cases} V_1(a) - V_0(b) - c & \text{if } R(a) \geq R(b) \\ V_0(a) - V_1(b) + c & \text{if } R(a) < R(b) \end{cases}. \tag{13}$$

The above lemma is summarized in Figure 2, in which the full-fledged four

panels describe $\Delta U(a, b; c)$ curves 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).

From the above lemma, we achieve two valuable propositions concerning 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). This implies that all the agents whose naturally oriented identity is favored in the labor market do not "switch", only some of those whose naturally oriented identity is less favored choose to "switch". 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: $\Sigma^A > (<) 0.5$ and $\Sigma^B < (>) 0.5$ if $a > (<) b$, as 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* $a > (<) b$.

Lemma 1 also 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. This implies that whenever $R(a) > R(b)$, the actual skill investment rate for the endogenously constructed group A (B) is greater (smaller) than that for the exogenously given group with the same collective reputation level: $\phi(a; b) > G(R(a))$ 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 the returns from the skill achievement are equal (i.e., $R(a) = R(b)$) even with $a \neq b$, $\Delta U(a, b; c)$ is constant with respect to c, and we have $\phi(a; b) = G(R(a)) = G(R(b)) = \phi(b; a)$.

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* $R(i) > R(j)$, $\phi(i;j) > G(R(i)) > G(R(j)) > \phi(j;i)$ *for each combination* $(i,j) \in \{(a,b),(b,a)\}$.[14]

For further analysis, we need to examine somewhat technical properties about the skill investment rates $\phi(a;b)$ resulting from the above noted 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 its 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 know that a dotted $\phi(a;b)$ curve must intercept the solid $\phi(a;a)(= G(R(a)))$ curve at both $a = b$ and $a = b'$, as described in the figure.

Next, it is very likely that the more attractive the choice of affect B is with the greater $R(b)$, the smaller the actual skill investment rate for the endogenously constructed group A, $\phi(a;b)$, will be. For instance, suppose that b is slightly greater than a, thus satisfying $a < b < \bar{\pi}$. Then, $\phi(a;b)$ will be slightly smaller than the actual skill investment rate without the endogenous group membership $G(R(a))$. As b increases, the return from skill investment $R(b)$ increases further (as far as $b < \bar{\pi}$), and the talented workers are more willing to take affect B, leading to the smaller $\phi(a;b)$ and the greater gap between $\phi(a;b)$ and $G(R(a))$. Reflecting this natural property, WLOG, we impose a *monotonicity condition* that for any specific $\tilde{b_1}$ and $\tilde{b_2}$, such that $\tilde{b_1} < \tilde{b_2} < \bar{\pi}$ (or $\bar{\pi} < \tilde{b_2} < \tilde{b_1}$), $\phi(a;\tilde{b_2})$

[14]This proposition implies that, given $R(i) > R(j)$ but $i < j$, it is even possible that the disproportionately less talented workers choose the 'affect' that corresponds to the favored employer belief j, resulting in $\phi(i;j) > \phi(j;i)$, for each combination $(i,j) \in \{(a,b),(b,a)\}$, as depicted in Panels B and C of Figure 2. This is because those who are talented gain less than those who are less talented with adopting the favored 'affect' j in such case: $V_1(j) - V_1(i) < V_0(j) - V_0(i)$.

is smaller than $\phi(a;\tilde{b_1})$ for each $a \in (0,1)$. This implies that the $\phi(a;\tilde{b_2})$ curve is placed below the $\phi(a;\tilde{b_1})$ curve, as seen in Figure 3. With the help of this property, we are able to avoid possible local fluctuations that might disturb our structural analysis of the global geometry.

Finally, 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*

$$\frac{\partial \phi(a;b)}{\partial a}\Big|_{a=b} \approx g(R(b))\,R'(b) + 2H'(0)\,R'(b)\,G(R(b)) \cdot [1 - G(R(b))]. \qquad (14)$$

Proof. Refer to the proof in the online 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), i.e., whenever b is less (greater) than $\bar{\pi}$. Additionally, the slope of $\phi(a;b)$ at the crossing point is greater (smaller) than the slope of $\phi(a;a)$, which equals $g(R(b))R'(b)$, whenever $R'(b)$ is positive (negative). These facts indicate that the slope of $\phi(a;b)$ is always "steeper" than $\phi(a;a)$ at such crossing point.

6 Characteristics of Endogenous Stereotyping

Now, we are ready to examine both the existence and the stability of Endogenous Stereotyping Equilibria. First of all, we show that allowing for endogenous group "switching" can increase the divergence in the reputation and actual skill acquisition rates across groups above the maximum divergence possible in a setting where there are multiple equilibria in the exogenous-groups case. WLOG, we assume that there exist three PSE 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 $\bar{\pi}$: $\pi_i < \bar{\pi}$, $\forall i \in \{l, m, h\}$.[15]

6.1 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. 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 the solid curve $\Gamma(b)^i$ in Figure 4 to see this unique notation rule.)

From the relative positions of the $\phi(x; y)$ curves for different levels of y, the properties of which are concretely discussed in the previous section, we can derive 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 $(\bar{\pi}, 1)$) (e.g. refer to the $\phi(a; b)$ curves and the relevant correspondences in Figure 3: given $b_1 < b_2 < \bar{\pi}$, $\Gamma(b_1)^h > \Gamma(b_2)^h$, $\Gamma(b_1)^l > \Gamma(b_2)^l$ and $\Gamma(b_1)^m < \Gamma(b_2)^m$). The curvature of the derived functions $\Gamma(y)^i$ 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 achieve both $\pi_h < \Gamma(0)^h < 1$ and $\pi_h < \Gamma(1)^h < 1$. Based on the above findings, the two correspondences $\Gamma(b)$, which is a set $\{a | a = \phi(a; b)\}$, and $\Gamma(a)$, which is a set $\{b | b = \phi(b; a)\}$, are depicted in solid and dashed curves for each and overlapped in each panel of Figure 4. Using the local linearization

[15] If there exist multiple PSE equilibria, two of them (denote by π_l and π_m) must be below $\bar{\pi}$ because it is assumed that $G(0) > 0$. (Refer to the equilibria in Panel B of Figure 1). Another possible one (denoted by π_h) can be greater or smaller than $\bar{\pi}$. We focus our analysis on the case with $\pi_h < \bar{\pi}$. However, readers will find that the main results do not change for the other possible case with $\pi_h > \bar{\pi}$, which are not presented in this manuscript but can be delivered upon request.

process, we can calculate the slope of *correspondence curve* at each trivial ESE $(\hat{x}, \hat{x})$, which satisfies $\hat{x} \in \Gamma(\hat{x})$, as follows:

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*

$$\Gamma'(\hat{x}) \approx \frac{2H'(0)\, R'(\hat{x})\, \hat{x}\, (1-\hat{x})}{g(R(\hat{x}))\, R'(\hat{x}) - 1 + 2H'(0)\, R'(\hat{x})\, \hat{x}\, (1-\hat{x})}. \tag{15}$$

Proof. Refer to the proof in the online appendix.■

Using the above lemma, we conclude that the slope of the *correspondence curve*, $\Gamma'(\hat{x})$, varies according to the density of the identity choice cost distribution around zero, $H'(0)$:

Lemma 5. *While 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,* $H'(0)$*:*

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

Proof. Based on the following three elementary facts, we can directly derive the results from Lemma 4: (1) $R'(\hat{x})$ is positive for any PSE $\hat{x}$ because we assume $\pi_i < \bar{\pi}, \forall i \in \{l, m, h\}$; (2) The slope of the $\phi(a; a)$ curve at $a = \pi_m$ is always greater than one: $g(R(\pi_m))R'(\pi_m) > 1$; (3) The slope of the $\phi(a; a)$ 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 these facts quickly from Figure 3.) ■

This lemma implies that when the sensitivity of identity choice activities represented by $H'(0)$ is sufficiently high that it is greater than some threshold

$\frac{1-g(R(\hat{x}))\,R'(\hat{x})}{4R'(\hat{x})\,\hat{x}\,(1-\hat{x})}$, 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.

The above lemmas help us to develop some meaningful theoretical conclusions. First, the following can be proved directly using the overlapped shapes of $\Gamma(a)$ and $\Gamma(b)$ in the (b,a) coordination plane:

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

Proof. Using Lemma 3, given multiple PSE (π_l,π_m and π_h) and the condition of $\pi_h<\bar{\pi}$, the *correspondence curve* $\Gamma(b)$ "passes through" the symmetric point (π_h,π_h) and a-intercept $(b,a)=(0,\Gamma(0)^h)$, in which $\pi_h<\Gamma(0)^h<1$. The *correspondence curve* $\Gamma(a)$ also "passes through" the symmetric point (π_l,π_l) and b-intercept $(a,b)=(1,\Gamma(1)^h)$, in which $\pi_h<\Gamma(1)^h<1$. (Refer to the panels of Figure 4.) Thus, there must 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 $b^*>a^*$. ■

In general, whether there are more than two ESE depends on the curvatures 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 $\hat{x}\in\{\pi_h,\pi_l\}$ generates two non-trivial ESE around the trivial ESE $(\hat{x},\hat{x})$, while the condition $|\Gamma'(\hat{x})|>1$ for $\hat{x}\in\{\pi_h,\pi_l\}$ does not generate such additional non-trivial ESE around the trivial ESE $(\hat{x},\hat{x})$. For instance, refer to Panel A of Figure 4 for the case with both $|\Gamma'(x_h)|<1$ and $|\Gamma'(x_l)|<1$ being satisfied, in which the total six non-trivial ESE are generated, and Panels B and C of the figure for the case with both $|\Gamma'(x_h)|>1$ and $|\Gamma'(x_l)|>1$ being satisfied, in which only two non-trivial ESE are generated.

Therefore, we can imagine the two non-trivial ESE that exist regardless of the curvatures of the correspondences $\Gamma(a)$ and $\Gamma(b)$. Let us call them "Persistent ESE" and denote them (π_L^*,π_H^*) and (π_H^*,π_L^*), in which both $\pi_H^*>\pi_h$ and

$\pi_L^* < \pi_l$ are satisfied as proved in the following theorem.

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. Given multiple PSE (π_l, π_m and π_h), we can find that $\pi_h < \Gamma(\pi_l)^h < \tilde{b}$, in which $\phi(\tilde{b}; \pi_l) = \pi_l$ as conjectured from Figure 3. Consequently, we know that $R(\Gamma(\pi_l)^h) > R(\tilde{b}) = R(\pi_l)$. Applying $i = \Gamma(\pi_l)^h$ and $j = \pi_l$ to Proposition 2, we obtain $G(R(\pi_l))(= \pi_l) > \phi(\pi_l; \Gamma(\pi_l)^h)$. This implies that $\Gamma(\Gamma(\pi_l)^h)^l < \pi_l$. (For instance, refer to the point $(\pi_l, \Gamma(\pi_l)^h)$ on the *correspondence curve* $\Gamma(b)^h$ and the point $(\Gamma(\Gamma(\pi_l)^h)^l, \Gamma(\pi_l)^h)$ on the *correspondence curve* $\Gamma(a)^l$ in Panel A of Figure 4.) Then, applying $\Gamma(\pi_l)^h > \pi_h$ and $\Gamma(\Gamma(\pi_l)^h)^l < \pi_l$, as $\Gamma(b)^h$ decreases over $b \in (0, \bar{\pi})$, there must be an intercept of the *correspondence curves* $\Gamma(b)$ 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^*). ■

The theorem implies that the inequality between endogenously constructed social groups in some non-trivial ESE can be greater than the inequality between exogenously given groups: e.g., $|\pi_H^* - \pi_L^*| > |\pi_h - \pi_l|$. This can occur because the former inequality is not only due to the positive complementarities between a group's reputation and its members' investment activities but also due to the positive selection along the ability parameter. That is, the group with the better collective reputation not only provides higher return to investment, but also attracts relatively more low-cost (high ability) workers from the disadvantaged group.

6.2 Stability of Endogenous Stereotyping Equilibria

For the examination of the stability of ESE, we consider the following 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

(rather than the affect B). 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. Right 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's 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.

In order to update their belief π_j, employers compare the realized actual skill acquisition rate of the entering newborns who adopt the affect j, $\phi(\pi_j; \pi_{-j})$, and their prior belief about the overall skill rate of the workers belonging to identity group j.[16] Whenever the realized skill acquisition rate of the newborns adopting the affect j, $\phi(\pi_j; \pi_{-j})$, is greater (smaller) than their prior belief about the skill level of identity group j, π_j, their posterior belief about the group's overall skill level becomes greater (smaller) than their prior one, as summarized in the following dynamics:

$$\dot{\pi}_j > (<)\, 0 \Leftrightarrow \phi(\pi_j; \pi_{-j}) > (<)\, \pi_j. \tag{16}$$

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 below $\Gamma(b)^l$ and between $\Gamma(b)^m$ and $\Gamma(b)^h$ 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 at the left-hand side of $\Gamma(a)^l$ and between $\Gamma(a)^m$ and $\Gamma(a)^h$ 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

[16] We assume that employers have correct information about the actual skill acquisition rate of the newborns belonging to each identity group.

following result.

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.*

This theorem together with Theorem 1 proves 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": $|\pi_H^* - \pi_L^*| > |\pi_i - \pi_j|$, $\forall i, j \in \{l, m, h\}$. When group membership is endogenous, 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.

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 know that $|\Gamma'(\hat{x})| > 1$ if and only if $H'(0) > \frac{1-g(R(\hat{x}))\,R'(\hat{x})}{4R'(\hat{x})\,\hat{x}\,(1-\hat{x})}$, 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 cannot be sustainable as any small perturbation would motivate a significant fraction of talented members to choose the "affect" associated with the better collective reputation, thereby leading to a divergence in the human capital cost distributions across groups that reinforces the disparity. Thus, we arrive at the following worthwhile result:

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

This means that when the affordability of identity choice activities is sufficiently high, the skill composition of the society inevitably converges to a unequal

"Persistent ESE" in the long run. From a policy perspective, this result provides a meaningful conclusion that, even with strong egalitarian government interventions, if the more talented individuals adopt the more highly regarded group's identity to a disproportionately very large extent, then the between-group difference will never be vanished. For instance, a change of accent (dialect) is one of the most affordable methods for regional identity manipulation. If talented members of a stereotyped regional group tend to modify their accents to avoid the anticipated disadvantages in the high-skilled labor market, the once-developed stereotypes against the group will not disappear, even when their government make strong commitments for establishing national unity and reconciliation between the groups.[17]

7 Welfare Properties and Implications

So far, we have provided an explicit micro-foundation for the endogenous group formation, which is embedded in a statistical discrimination framework with endogenous beliefs about skill acquisition. This allows for some welfare analysis, highlighting the winners and losers from the assimilation process. Among many situations in which identity choice and stereotypes operate in tandem, we focus on the welfare effects resulting from the two identity manipulation activities introduced in Section 2: passing and 'partial passing' behaviors.

7.1 Selective Out-migration (Passing)

Consider two social groups, a privileged group (A) and a stigmatized group (B). The selective out-migration from the stigmatized group to the privileged group occurs when the return for "passing" such as better treatment in the

[17]In facing these challenges, a government may consider the policies that may affect the sensitivity of identity choices (as captured by $H'(0)$), such as group specific (religious, ethnic, cultural or regional) educational programs or public promotion of social events celebrating specific identity categorizations.

labor market outweighs its cost such as losing ties to ones' own kind, learning unfamiliar customs and adopting a new culture. According to Theorem 2, 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 perceived identity is not malleable and each group's collective reputation is self-confirmed at one of the stable PSE: π_h for the privileged group and π_l for the stigmatized group.[18] Refer to Panel A of Figure 5 for this benchmark economy without the passing activities.[19]

Now, let us clarify who benefits and who suffers from the prevalence 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 being stigmatized in the marketplace ($k \geq \Delta U(\pi_H^*, \pi_L^*; c)$) and "the advantaged" who keep their privileged type-A membership ($k \leq 0$).

Because the anticipated net reword $U(\pi, c)$ is monotonically increasing in π and the condition $\pi_L^* < \pi_l < \pi_h < \pi_H^*$ holds according to Theorem 1, we can infer that "non-passers" suffer from the prevalent out-migration activities as much as $U(\pi_l, c) - U(\pi_L^*, c)$, while "the advantaged" benefit from such activities as much as $U(\pi_H^*, c) - U(\pi_h, c)$. It is noteworthy that not all of passers benefit from the prevalence 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

[18]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.

[19]In the theoretical model, the identity manipulation cost k is symmetrically distributed around zero. In the real world, however, we see that many stereotyped groups are in fact minorities. Acknowledging this reality does not make a qualitative difference in terms of model interpretations. The only difference is that the decline in the reputation of the minority group is affected more by existing passing activities, while the increased reputation of the dominant group is less affected by them.

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 $\tilde{k}(c) \equiv U(\pi_H^*, c) - U(\pi_l, c)$.

Note that the threshold $\tilde{k}(c)$ satisfies $0 < \tilde{k}(c) < \Delta U(\pi_H^*, \pi_L^*; c)$ for any specific level of c. Then, we achieve the following welfare property that denies the possibility of *Parato improvement*.

Proposition 5. *The individuals (with skill investment cost c) whose identity manipulation cost k is above the threshold $\tilde{k}(c)$ suffer due to the prevalence of passing activities, while those whose identity manipulation cost k is below the threshold benefit from it.*

Second, let us examine the conditions under which the selective out-migration may improve the social efficiency. We can compute the societal efficiency gain (ΔW_{total}) by the double integrations of the welfare changes of the three population aggregates (non-passers, passers and the advantaged):[20]

$$\begin{aligned}\Delta W_{total} &= \int_0^\infty \Big[\int_{\Delta U}^\infty [U(\pi_L^*, c) - U(\pi_l, c)]\, \mathrm{d}H(k) + \int_0^{\Delta U} [U(\pi_H^*, c) - U(\pi_l, c) - k]\, \mathrm{d}H(k) \\ &\quad + \int_{-\infty}^0 [U(\pi_H^*, c) - U(\pi_h, c)]\, \mathrm{d}H(k) \Big]\, \mathrm{d}G(c), \text{ where } \Delta U \equiv \Delta U(\pi_H^*, \pi_L^*; c)\end{aligned}$$

Through the decomposition, we obtain[21]

$$\begin{aligned}\Delta W_{total} &= \underbrace{0.5 \int_0^\infty [U(\pi_H^*, c) - U(\pi_h, c)]\, \mathrm{d}G(c)}_{\text{“positive reputational externality”}} - \underbrace{0.5 \int_0^\infty [U(\pi_l, c) - U(\pi_L^*, c)]\, \mathrm{d}G(c)}_{\text{“negative reputational externality”}} \\ &\quad + \underbrace{\int_0^\infty \int_0^{\Delta U} [\Delta U - k]\, \mathrm{d}H(k)\, \mathrm{d}G(c)}_{\text{“passing premium”}}, \text{ using the symmetry of } H(k). \qquad (17)\end{aligned}$$

[20] The employers' expected payoffs are always zero because they are assumed to pay exact wages to workers according to their expected productivity.

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

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. Both externalities are summarized in the first and second terms in the above equation.[22] The third term in the equation plays a significant role in the determination of the positive societal efficiency gain, which reflects the passing premium for the passers who choose to elect type A although their natural orientation is type B. The 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 minus the size of the positive reputational externality. Therefore, the above decomposition delivers the following welfare implication:

Proposition 6. *The selective out-migration behaviors can cure to some extent the social inefficiency caused by the labor market imperfection as far as 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 big enough that it surpasses the net loss in terms of the reputational externalities.*[23]

Using the symmetry assumption of $H(k)$, 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

[22]However, the advantaged group, besides being concerned about returns to skills, may care more about maintaining its social status position. If this be the case, the members of the advantaged group would create "subtle" socialization barriers to members of the other groups, making it more difficult for them to "pass"; this brings to mind sociologist Bourdieu's (1987) term "distinction."

[23]In the given model, the wage rate per unit of efficient high-skilled (low-skilled) labor is fixed as w (0). However, if we allow the skill complementarities between high and low skill labor in production, the wage rate per unit of high-skilled (low-skilled) labor would depend negatively (positively) on the total level of human capital in the economy. Since the selective out-migration tends to raise the total level of human capital, this would reduce the benefits of the "passing premium" and the size of the positive reputational externality as well as the size of the negative reputational externality. The societal efficiency gain of passing would then be reduced.

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 (ΔW_{total}) will be. Accordingly, the positive efficiency gain is more likely to be achieved when identity manipulation is easier to undertake.

It is also notable that 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 $\pi_L^* < \pi_l\,(\approx 0)$. That is, in this extreme case, there is no reputation to lose for the disadvantaged group members, at least not as a result of the endogenous out-migration. Therefore, it is *Pareto-improving*, in that a positive societal efficiency gain is achieved with "non-passers" who are at least as well off, and all other agents who are better off:

Proposition 7. *The selective out-migration from a severely stigmatized group is Pareto-improving without hurting the welfare of the left-behind.*

There are many real-life examples in which passing improves social efficiency. For instance, the living conditions of the Zainichi were the worst in Japan, and they were severely stigmatized even after Japanese imperialism ended. However, their identity manipulation was relatively easier to achieve, given how their appearance was similar to that of Japanese individuals. Their selective out-migration presumably improved social efficiency: the passing premium was high, but the negative impact on those left behind was minimal. Given the severe discrimination against the Zainichi, it could even be *Pareto-improving*.

7.2 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). Most of the better-off members of this stigmatized population will not be able 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.[24] In this way, a "visible" subgroup can be constructed around any cluster of markers which are evidently informative though functionally irrelevant traits (such as affectations of speech, dressing up and consumption habits).[25]

Imagine a specific set of indices that is used for the differentiation. Suppose that employers, who are doing their best under trying circumstances, 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 not to adopt the indices ($k > 0$) and the other subpopulation whose natural orientation is to adopt the indices ($k < 0$): an agent with positive k should incur the cost k to be equipped with those indices, while an agent with negative k should incur the cost $-k$ to discard their naturally adopted indices.

The theory developed in this paper is directly applied to this altered setting, replacing groups A and B with subgroups Z' and Z. The most talented members of the population, who gain most by separating themselves from the mass, 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

[24] By using a more refined set of indices to guide their discrimination, observers may also encourage the production of those very indices of differentiation by the more talented members.

[25] Anything that is costly to acquire can be one of the markers, but the most effective ones to send signals that they are different will be cultural or behavioral traits of the better regarded social group, because successfully adopting those traits will signal a person's willingness to put in effort to "confirm", which is valuable to employers.

which $\pi_L^* < \pi_l < \pi_h < \pi_H^*$ holds.

The welfare effects of this 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 in the stigmatized group do not make a strategic decision on whether to adopt the indices. In this benchmark economy, there should be no clear difference in terms of the skill acquisition rates between the two subgroup Z and Z': $(z, z') = (\pi_l, \pi_l)$. (Refer to Panel B of Figure 5 for this benchmark economy.) Then, we obtain the following welfare changes of three population aggregates:

Lemma 6. *Comparing an unequal stable economy with the prevalent 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 prevalence of the activities as much as $U(\pi_l, c) - U(\pi_L^*, c)$, while the population who adopt the indices naturally $(k < 0)$ is benefited from it as much as $U(\pi_H^*, c) - U(\pi_l, c)$. The welfare change of a "partial passer $(0 < k < \Delta U(\pi_H^*, \pi_L^*; c))$" is $U(\pi_H^*, c) - U(\pi_l, c) - k$, which is positive (negative) for those whose cost to adopt the indices is less (greater) than the threshold $\tilde{k}(c)(\equiv U(\pi_H^*, c) - U(\pi_l, c))$, which satisfies $0 < \tilde{k}(c) < \Delta U(\pi_H^*, \pi_L^*; c), \forall c$.*

The welfare results could help shed light on the conflict within a stereotyped population. While some partial passers whose identity manipulation cost is lower benefit from those activities, the non-partial passers suffer from them. The worse-off members of the group may accuse the partial passers of some kind of immoral betrayal, which is often referred to as "acting white" in the US racial context due to the partial passers' assuming the social expectations of white society.[26] Thus, social identity manipulation through partial passing can

[26]For instance, the behaviors that lead to accusation of "acting white" include speaking standard English, wearing clothes from the Gap or Abercrombie & Fitch, wearing shorts in the winter, and enrolling in honors or advanced placement classes, according to Neal-Barnett's (2001) focus group interview. Among them, academic success is a functionally relevant index, which is valuable to employers. Thus, it will further exaggerate the positive selection effects.

be a way to undermine solidarity in the visibly distinct stigmatized population. The adverse impact on the left-behind may generate the resentment against the partial passers. Furthermore, the worse-off members may try to hold such people back by stigmatizing their action to adopt the indices of differentiation.[27]

This is an alternative explanation of the "acting white" phenomenon to that offered by Austen-Smith and Fryer (2005). They propose 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.[28]

The societal efficiency gain from the prevalence of partial passing activities is computed by the double integrations of the welfare changes summarized in Lemma 6, or, alternatively, by the replacement of $U(\pi_h, c)$ with $U(\pi_l, c)$ in ΔW_{total} (equation (17)). In general, Propositions 5 and 6 work for this altered setting: the positive efficiency gain is achieved only when the premium obtained by the partial passers is greater than the net loss in terms of the reputational externalities. The positive efficiency gain is more likely to be achieved when the adoption of the indices of differentiation is easier to undertake.[29]

Applying this thinking to the "acting white" phenomenon, we find that the supposed "immoral" activities of which some are accused may improve the total welfare of society, though they may engender some conflicts within the population. The improvement is clear when the adoption of those indices are not very

[27]Moreover, the negative impact on those left behind may trigger internal reactions that have the intended goal of increasing the identity manipulation cost. This may explain the emergence of collective institutions (e.g., gangs, religious or ethnic associations) within the disadvantaged group that will try to shift the distribution of the cost k.

[28]In a similar spirit, scholars in Sociology (e.g., Wilson, 1987) argue that the movement of the black middle-class from black neighborhoods to suburbs (so-called "black flight") has had a detrimental impact on black poverty.

[29]The partial passing premium is identical to the passing premium in equation (17).

costly and the disadvantaged population had been widely stigmatized in society, because then the premium obtained by partial passers will be great, but the size of the created negative reputational externalities will be relatively smaller (cf. Proposition 7).

8 Conclusion

Our theoretical model 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, 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 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 similar inequality-amplifying effects of heterogeneous incentives for mobility are also found in other areas of the inequality literature, such as that on school vouchers, which reduces the switching costs for bright kids in moving from poor public schools to affluent private ones (Epple and Romano, 1998); socioeconomic stratification in a city, which arises due to middle-class flight to the suburbs (Benabou, 1993); and brain drain, which is caused by immigrant self-selection (Borjas, 1987). In these cases, rather than social identity, the better-off types (i.e., those of high income or high ability) choose their school, neighborhood, or country without taking into account the effects of their choice on others.

Our theoretical model is a step toward a better understating of various identity choice behaviors. We have applied the theory to the passing and 'partial passing' phenomena, finding that these inequality-amplifying 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. 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. One might expect the winners and losers to take actions accordingly—namely, the punishing activity by the losers, to deter selective out-migration, and the subsidies offered by the winners, to promote it. The government may consider policy measures that are more likely to mitigate (or amplify) the reinforcing effects between endogenous identity and investment incentives. These reactions of the stakeholders and their economic implications are worthy of further examination, but are left for future study.

The developed micro-foundation of endogenous group formation has the potential to illuminate many other social phenomena involving the choice of the perceived identities (e.g., racial profiling in law enforcement, the coming out decision by LGBT people and effectively 'branding' a new consumer product). In the increasingly globalized and multicultural societies in which we live, the question of identity choice and how this interacts with investment incentives and socioeconomic discrimination processes is becoming an important topic with many policy implications. Thus, we look forward to seeing more developments in the economic research on this topic of endogenous identity choice.

References

Akerlof, George and Rachel Kranton. "Economics and Identity." *The Quarterly Journal of Economics*, 2000, 115(3), pp. 716–53.

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.

Austen-Smith David and Roland Fryer. "An Economic Analysis of 'Acting White'." *The Quarterly Journal of Economics*, 2005, 36(2), pp. 551–83.

Benabou, Roland. "Workings of a City: Location, Education, and Production." *The Quarterly Journal of Economics*, 1993, 108(3), pp. 619-52.

Benabou, Roland and Jean Tirole. "Identity, Morals, and Taboos: Beliefs as Assets." *The Quarterly Journal of Economics*, 2011, 126, pp. 805–55.

Benjamin, Daniel, James Choi, and A. Joshua Strickland. "Social Identity and Preferences." *American Economic Review*, 2010, 100(4), pp. 1913–28.

Borjas, George. "Self-Selection and the Earnings of Immigrants." *American Economic Review*, 1987, 77(4), pp. 531–53.

Bourdieu, Pierre. *Distinction: A Social Critique of the Judgement of Taste.* Harvard University Press, 1987.

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.

Eppel, Dennis and Richard Romano. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effect." *American Economic Review*, 1998, 88(1), pp. 33–62.

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

Fang, Hanming and Glenn Loury. " 'Dysfunctional Identities' Can be Rational." *American Economic Review*, 2005, 95, pp. 104–11.

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.

Hunt, Earl. *Human Intelligence*. Cambridge University Press, 2011.

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.

Neal-Barnett, Angela. "Being Black: A New Conceptualization of Acting White." In A. NealBarnett, J. Contreras, and K. Kerns (Eds.), *Forging Links: African American Children Clinical Developmental Perspectives*. Westport, CT: Greenwood Publishing Group, 2001.

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.

Wilson, William. *The Truly Disadvantaged: The Inner City, the Underclass and*

Public Policy. Chicago: University of Chicago Press. 1987.

[For Online Publication: Appendixes A and B]

Appendix A: ESE in a Simple Model

In this appendix, we present the endogenous stereotyping equilibria in the simplest possible cost and signaling structure, in order to help readers better understand the elementary mechanism of endogenous group formation at work.

A1. The Simplest Cost and Signaling Structure

Let us adopt discrete cost and labor market signal distributions, instead of continuous ones. The population comprises three human capital investment cost (c) types: (1) Π_l fraction of agents whose investment cost (C_l) is close to zero and who will thus always invest in skills, (2) $\Pi_h - \Pi_l$ fraction of agents whose investment cost (C_m) is mediocre, and who will decide whether or not to invest based on the expected return to skill investment, and (3) $1 - \Pi_h$ fraction of agents whose investment cost (C_h) is very high and who will never invest in skills. Then, we have the step function of $G(c)$: $G(c) = \Pi_l, \forall c \in (0, C_m)$; $G(c) = \Pi_h, \forall c \in [C_m, C_h)$; $G(c) = 1, \forall c \in [C_h, \infty]$. In terms of the relative cost of being perceived as A rather than B (k), the population comprises four types: η fraction (η fraction) of agents who are naturally inclined toward B (A) and should incur a relatively lower cost K_l to be perceived as A (B), indicating that $k = K_l$ ($k = -K_l$), and $0.5 - \eta$ fraction ($0.5 - \eta$ fraction) of agents who are naturally inclined toward B (A) and should incur a very high cost K_h to be perceived as A (B), indicating that $k = K_h$ ($k = -K_h$). Thus, we have in total 12 different population aggregates, of which the cost combination (c, k) is represented by $(c, k) \in \{(C_i, K_j), (C_i, -K_j)\}, \forall i \in \{l, m, h\}, \forall j \in \{l, h\}$. (Refer to the distribution of those 12 aggregates, as seen in Panels A and B of Appendix Figure 1.)

The test of qualification (prior to assignment) yields one of the three signals, $t \in \{H, M, L\}$. The test outcome H (L) is achieved only by those who are

qualified (unqualified). The test outcome M can be achieved by either those who are qualified or those who are unqualified. Let P_q (P_u) be the probability that if a worker does (does not) invest, his or her test outcome is M: $P_q \equiv Prob[M|skilled]$ and $P_u \equiv Prob[M|unskilled]$.

We further assume that workers receive a gross benefit of W if hired, and zero if unemployed. Employers gain a net return of X_q if they hire a skilled worker, and suffer a net loss of X_u if they hire an unskilled worker. Then, they will (will not) hire all who achieve signal H (L), and will hire a worker who achieves signal M if and only if the expected net return from doing so is nonnegative: $X_q \cdot Prob[skilled|M] - X_u \cdot Prob[unskilled|M] \geq 0$, in which the posterior probability that the worker with the unclear test outcome M is in fact skilled is $Prob[skilled|M] = \pi P_q/(\pi P_q + (1-\pi)P_u)$, using Bayes' rule, and given the believed skill investment rate of the group π. Hence, employers will hire a worker with signal M if and only if the employer is sufficiently optimistic about the rate of skill acquisition of a group from which the worker was drawn:

$$\textit{Hiring a worker with signal } M \iff \pi \geq \frac{P_u X_u}{P_q X_q + P_u X_u} (\equiv \Pi^*), \tag{18}$$

for which we assume that the threshold level Π^* satisfies $\Pi_l < \Pi^* < \Pi_h$.

Given this simplest framework, the expected payoff from acquiring a skill $V_1(\pi)$ is W if $\pi \geq \Pi^*$, and $W(1-P_q)$ if $\pi < \Pi^*$. That without acquiring a skill $V_0(\pi)$ is WP_u if $\pi \geq \Pi^*$, and 0 if $\pi < \Pi^*$. Thus, the expected economic return from being skilled $R(\pi)(\equiv V_1(\pi) - V_0(\pi))$ is

$$R(\pi) = \begin{cases} W(1-P_u), & \text{if } \pi \geq \Pi^* \\ W(1-P_q), & \text{if } \pi < \Pi^* \end{cases}. \tag{19}$$

In order to have multiple PSEs, Π_l and Π_h, the human capital investment costs must satisfy the condition $C_l \leq W(1-P_q) < C_m \leq W(1-P_u) < C_h$, because $G(R(\Pi_l)) = G(W(1-P_q)) = \Pi_l$ only when $C_l \leq W(1-P_q) < C_m$, and

$G(R(\Pi_h)) = G(W(1-P_u)) = \Pi_h$ only when $C_m \leq W(1-P_u) < C_h$.

A2. ESEs Given Multiple PSEs (Π_l and Π_h)

Now suppose that perceived identity is malleable and groups are endogenously constructed, given the existence of multiple PSEs, Π_l and Π_h. The anticipated net reward for a worker who belongs to a group believed to be investing at rate π, denoted by $U(\pi, c)$, is either $V_1(\pi) - c$ if he or she invests, or $V_0(\pi)$ if he or she does not. Hence, it is expressed as $\max\{V_1(\pi) - c, V_0(\pi)\}$:

$$U(\pi, c) = \begin{cases} \max\{W - c, WP_u\}, & \text{if } \pi \geq \Pi^* \\ \max\{W(1-P_q) - c, 0\}, & \text{if } \pi < \Pi^* \end{cases}. \tag{20}$$

Given the employers' prior belief about human capital investment rates (a, b), we achieve a worker's incentive for electing type-A rather than type-B, denoted by $\Delta U(a, b; c)$, which is equivalent to $U(a, c) - U(b, c)$. Only the population aggregate whose cost set (c, k) satisfies $k \leq \Delta U(a, b_1; c)$ will elect type-A. All the other aggregates will elect type-B.

When both a and b are less (or greater) than Π^*, this incentive is zero, indicating that those whose k is negative (positive) elect type-A (type-B). Given $b < \Pi^* < a$, however, this incentive is positive for every human capital cost type and non-increasing in c, as seen in Panel A of Appendix Figure 1:

$$\Delta U(a, b; c) = \begin{cases} WP_q, & \text{if } c = C_l \\ W - C_m, & \text{if } c = C_m \\ WP_u, & \text{if } c = C_h \end{cases}, \text{ given } b < \Pi^* < a. \tag{21}$$

In a symmetrical manner, given $a < \Pi^* < b$, this incentive is negative for every human capital cost type and non-decreasing in c, as seen in Panel B of the same

figure:

$$\Delta U(a,b;c)=\begin{cases}-WP_q, & \text{if } c=C_l\\ -W+C_m, & \text{if } c=C_m\ ,\ \text{given } a<\Pi^*<b.\\ -WP_u, & \text{if } c=C_h\end{cases} \tag{22}$$

Before we search for ESEs, we further impose for the sake of simplicity that K_h is sufficiently high that $K_h > WP_q$, while K_l is greater than $W - C_m$ but smaller than WP_q: $W-C_m < K_l < WP_q < K_h$. Then, when group B's believed investment rate is assumed to be less than the threshold Π^* (i.e., $b_1 < \Pi^*$), the actual skill investment rate for the endogenously constructed group A, denoted by $\phi(a;b_1)$, is $\Pi_l, \forall a \in [0,\Pi^*)$, because $\Delta U(a,b_1;c)=0$ and $R(a)=W(1-P_q)<C_m$. It is $\Pi'_h, \forall a \in [\Pi^*,1]$, in which $\Pi'_h = (0.5\Pi_h+\Pi_l\eta)/(0.5+\Pi_l\eta)\,(>\Pi_h)$, because only those whose cost set is (C_l,K_l) will "switch" from his or her own natural orientation B to type-A and $R(a)=W(1-P_u)\geq C_m$, as seen in Panel A of the figure.

On the other hand, when group B's believed investment rate is assumed to be greater than the threshold (i.e., $b_2 > \Pi^*$), the actual skill investment rate for the endogenously constructed group A, $\phi(a;b_2)$, is $\Pi'_l, \forall a \in [0,\Pi^*)$, in which $\Pi'_l = (0.5\Pi_l-\Pi_l\eta)/(0.5-\Pi_l\eta)\,(<\Pi_l)$, because only the population aggregate with its cost set $(C_l,-K_l)$ will "switch" from its natural orientation A to type-B and $R(a)=W(1-P_q)<C_m$, as noted in Panel B of the figure. It is $\Pi_h, \forall a \in [\Pi^*,1]$, because $\Delta U(a,b_2;c)=0$ and $R(a)=W(1-P_u)\geq C_m$. Hence, we achieve the step functions of $\phi(a;b_1)$ and $\phi(a;b_2)$, which are depicted in Panel C of the figure, together with their benchmark curve $\phi(a;a)$, in which the believed investment rates for the two groups are equal: $\phi(a;a)$ is $\Pi_l, \forall a \in [0,\Pi^*)$, and $\Pi_h, \forall a \in [\Pi^*,1]$.

Using these actual skill investment rate functions $\phi(a;b)$, we can compute the correspondence $\Gamma(b)$, which is a set of group A's believed skill investment rates

that are self-confirmed by its actual skill investment rates, given that the other group B's believed skill investment rate is fixed as b: $\Gamma(b) = \{a|a = \phi(a; b)\}$. From the functions $\phi(a; b_1)$ and $\phi(a; b_2)$ and a 45-degree line in Panel C, we infer that when $b < \Pi^*$, $\Gamma(b) = \{\Pi_l, \Pi'_h\}$, while $\Gamma(b) = \{\Pi'_l, \Pi_h\}$ when $b \geq \Pi^*$. By symmetry, we also have $\Gamma(a) = \{\Pi_l, \Pi'_h\}$ when $a < \Pi^*$, and $\Gamma(a) = \{\Pi'_l, \Pi_h\}$ when $a \geq \Pi^*$. A set of ESEs (Ω_{KL}) is a set of (a, b)s that satisfy both $a \in \Gamma(b)$ and $b \in \Gamma(a)$. Using the two correspondences $\Gamma(b)$ and $\Gamma(a)$ overlapped in Panel D, we can identify four ESEs: two trivial ESEs, (Π_l, Π_l) and (Π_h, Π_h), and two nontrivial ESEs, (Π'_l, Π'_h) and (Π'_h, Π'_l). Thus, knowing that $\Pi'_h > \Pi_h$ and $\Pi'_l < \Pi_l$, we prove that the inequality between endogenously constructed social groups can be greater than the inequality that can emerge between exogenously given groups: $|\Pi'_h - \Pi'_l| > |\Pi_h - \Pi_l|$.

Appendix B: Proofs

Proof of Lemma 2:

Consider a very small $\delta > 0$ such that $a = b + \delta$. We can denote $\sigma^A(\delta; b)$ and $\Sigma^A(\delta; b)$, which are functions of δ given b, and consequently $\sigma^{A'}(\delta; b)$ and $\Sigma^{A'}(\delta; b)$, which are the corresponding partial derivatives with respect to δ. The slope of the $\phi(a; b)$ curve at a=b can be expressed as follows, using $\sigma^A(\delta; b)$ and $\Sigma^A(\delta; b)$,

$$\begin{aligned}
\frac{\partial \phi(a;b)}{\partial a}\Big|_{a=b} &= \lim_{\delta \to 0} \frac{\phi(b+\delta;b) - \phi(b;b)}{\delta} \\
&= \lim_{\delta \to 0} \frac{\sigma^A(\delta;b)/\Sigma^A(\delta;b) - \sigma^A(0;b)/\Sigma^A(0;b)}{\delta} \\
&= \lim_{\delta \to 0} \left[\frac{[\sigma^A(\delta;b) - \sigma^A(0;b)]\Sigma^A(0;b)}{\delta} - \frac{[\Sigma^A(\delta;b) - \Sigma^A(0;b)]\sigma^A(0;b)}{\delta} \right] \\
&\quad \cdot \frac{1}{\Sigma^A(\delta;b) \cdot \Sigma^A(0;b)} \\
&= \frac{\sigma^{A'}(0;b) \cdot \Sigma^A(0;b) - \sigma^A(0;b) \cdot \Sigma^{A'}(0;b)}{\lim_{\delta \to 0} \Sigma^A(\delta;b) \cdot \Sigma^A(0;b)} \qquad (23)
\end{aligned}$$

To compute this outcome, first, 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 also know $H'(k) \approx H'(0)$ for small enough k. Then, using Lemma 1 and Panels A and B of Figure 2, the fraction of agents who elect to be A-type and decide to be skilled, $\sigma^A(\delta; b)$, and its derivative, $\sigma^{A'}(\delta; b)$, are approximated by

$$\begin{aligned}
\sigma^A(\delta;b) &\approx G(R(b) + \Delta) \cdot [0.5 + H'(0)\,(V_1(b+\delta) - V_1(b))] - 0.5\,H'(0)\,g(R(b))\,\Delta^2, \\
\sigma^{A'}(\delta;b) &\approx g(R(b) + \Delta)\,R'(b+\delta)[0.5 + H'(0)\,(V_1(b+\delta) - V_1(b))] \\
&\quad + G(R(b) + \Delta)\,H'(0)\,V_1'(b+\delta) - H'(0)\,g(R(b))\,\Delta\,R'(b+\delta),
\end{aligned}$$

in which the last terms that are related to a triangle area in the figure, $-0.5\,H'(0)\,R'(b)\,\Delta^2$ 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). Similarly, the fraction of agents who

elect to be A-type, $\Sigma^A(\delta; b)$, and its derivative, $\Sigma^{A'}(\delta; b)$, are approximated by

$$\begin{aligned}
\Sigma^A(\delta; b) &\approx 0.5 + H'(0)\,(V_0(b+\delta) - V_0(b)) + G(R(b)+\Delta)\,H'(0)\,\Delta - 0.5\,H'(0)\,g(R(b))\,\Delta^2, \\
\Sigma^{A'}(\delta; b) &\approx H'(0)\,V_0'(b+\delta) + G(R(b)+\Delta)\,H'(0)\,R'(b+\delta) + g(R(b)+\Delta)\,R'(b+\delta)\,H'(0)\,\Delta \\
&\quad -H'(0)\,g(R(b))\,\Delta\,R'(b+\delta).
\end{aligned}$$

Using the above approximations, we achieve the following results when $\delta = 0$:

$$\begin{cases}
\sigma^A(0; b) & \approx 0.5\,G(R(b)) \\
\sigma^{A'}(0; b) & \approx 0.5\,g(R(b))\,R'(b) + G(R(b))\,H'(0)\,V_1'(b) \\
\Sigma^A(0; b) & \approx 0.5 \\
\Sigma^{A'}(0; b) & \approx H'(0)\,V_0'(b) + G(R(b))\,H'(0)\,R'(b)
\end{cases} \tag{24}$$

Applying these results and $\lim_{\delta\to 0}\Sigma^A(\delta; b) = 0.5$ to equation (23), we have the following approximation, noting $R'(b) \equiv V_1'(b) - V_0'(b)$:

$$\frac{\partial\phi(a; b)}{\partial a}\Big|_{a=b} \approx g(R(b))\,R'(b) + 2\,H'(0)\,R'(b)\,G(R(b))\cdot[1 - G(R(b))]. \tag{25}$$

QED.

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 2. 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 $(\hat{x}, \hat{x})$, which equals $g(R(\hat{x}))R'(\hat{x})$, the correspondence value x' approximately satisfies the following condition, as conjectured from the figure:

$$x' - [\hat{x} + g(R(\hat{x}))\,R'(\hat{x})\,\Delta] \approx \frac{\partial\phi(x; y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta}\cdot[x' - (\hat{x} + \Delta)]\,. \tag{26}$$

The slope of the *correspondence curve* at a trivial ESE $(\hat{x}, \hat{x})$, denoted by $\Gamma'(\hat{\pi})$,

is approximately equal to $\lim_{\Delta \to 0} \frac{x' - \hat{x}}{\Delta}$:

$$\Gamma'(\hat{x}) \approx \left[g(R(\hat{x}))\, R'(\hat{x}) - \lim_{\Delta \to 0} \frac{\partial \phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta} \right] \Big/ \left[1 - \lim_{\Delta \to 0} \frac{\partial \phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta} \right]. \tag{27}$$

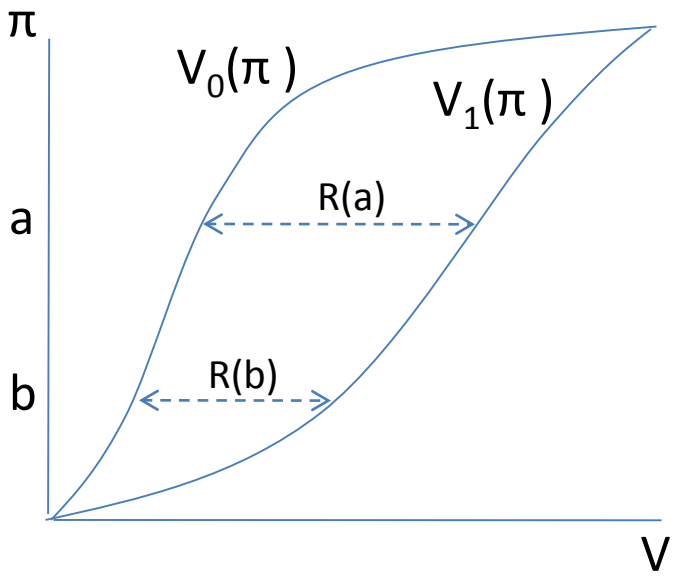
From Lemma 2 and $G(R(\hat{x})) = \hat{x}$, we have

$$\lim_{\Delta \to 0} \frac{\partial \phi(x;y)}{\partial x}\Big|_{x=y=\hat{x}+\Delta} = g(R(\hat{x}))R'(\hat{x}) + 2H'(0)R'(\hat{x})\hat{x}(1-\hat{x}). \tag{28}$$

Applying this result to equation (27), we achieve the given result for $\Gamma'(\hat{x})$. The examples for the positive $\Gamma'(\hat{x})$ and the negative $\Gamma'(\hat{x})$ are depicted separately in Panels A and B of Appendix Figure 2. QED.

Figure 1. Phenotypic Stereotyping Equilibria

Panel A. Expected Payoffs Given π



Panel B. Multiplicity of Equilibria

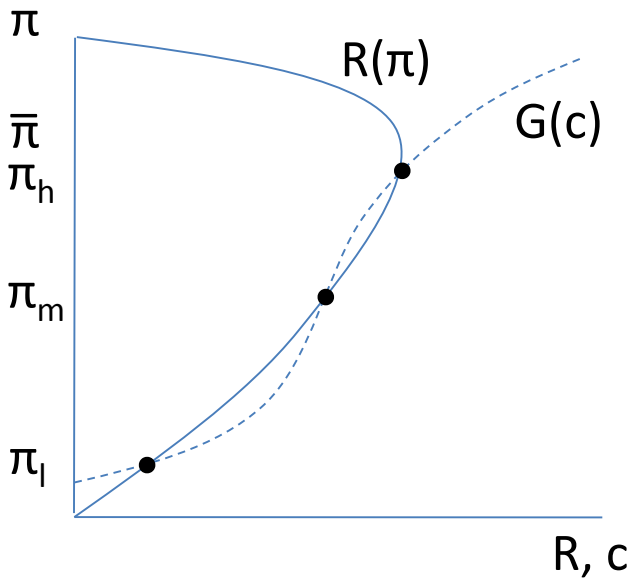
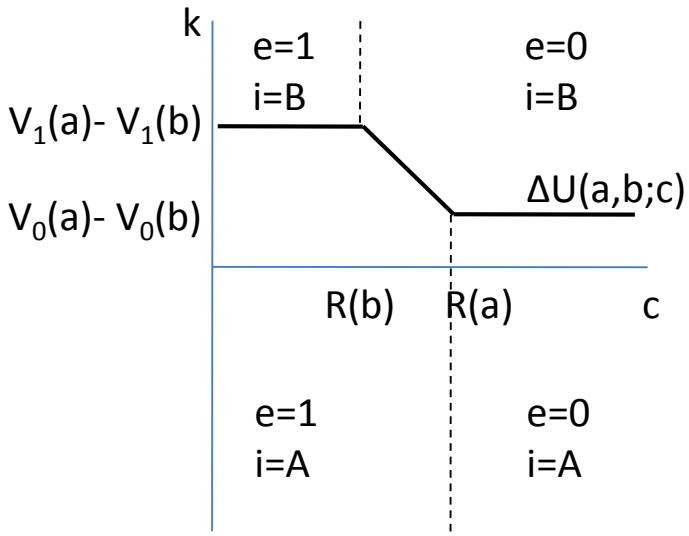
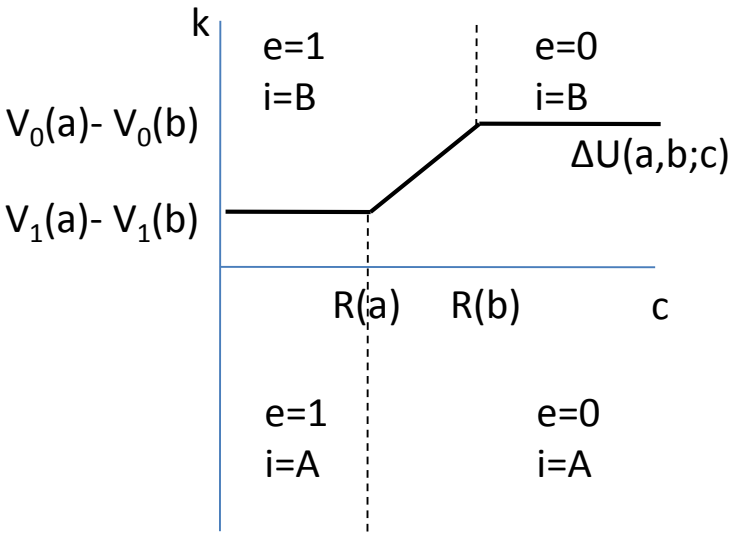


Figure 2. Human Capital Investment and Identity Choice Behavior

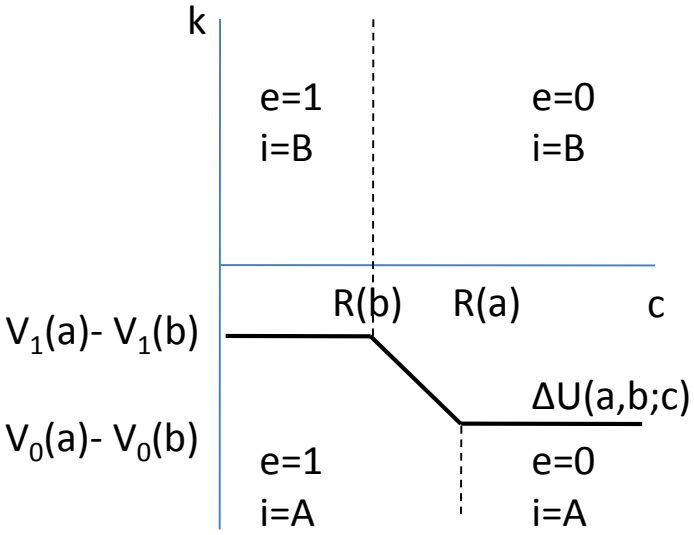
Panel A. Case with a>b and R(a)>R(b)



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)

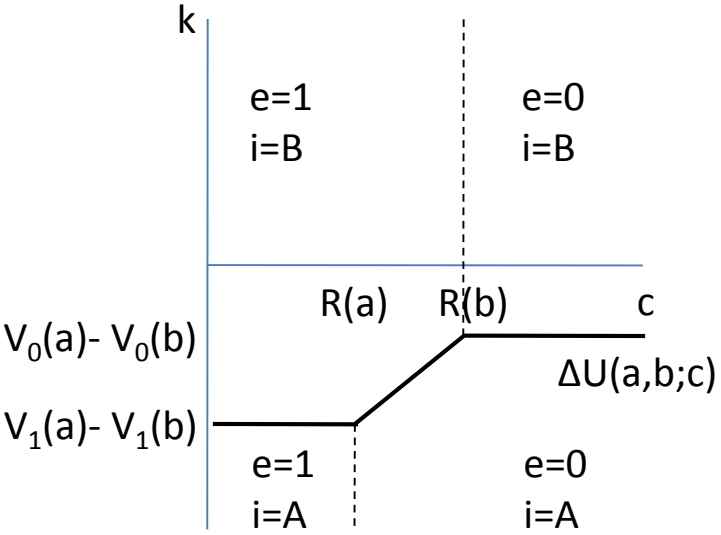


Figure 3. Human Capital Investment Rate ϕ(a; b)

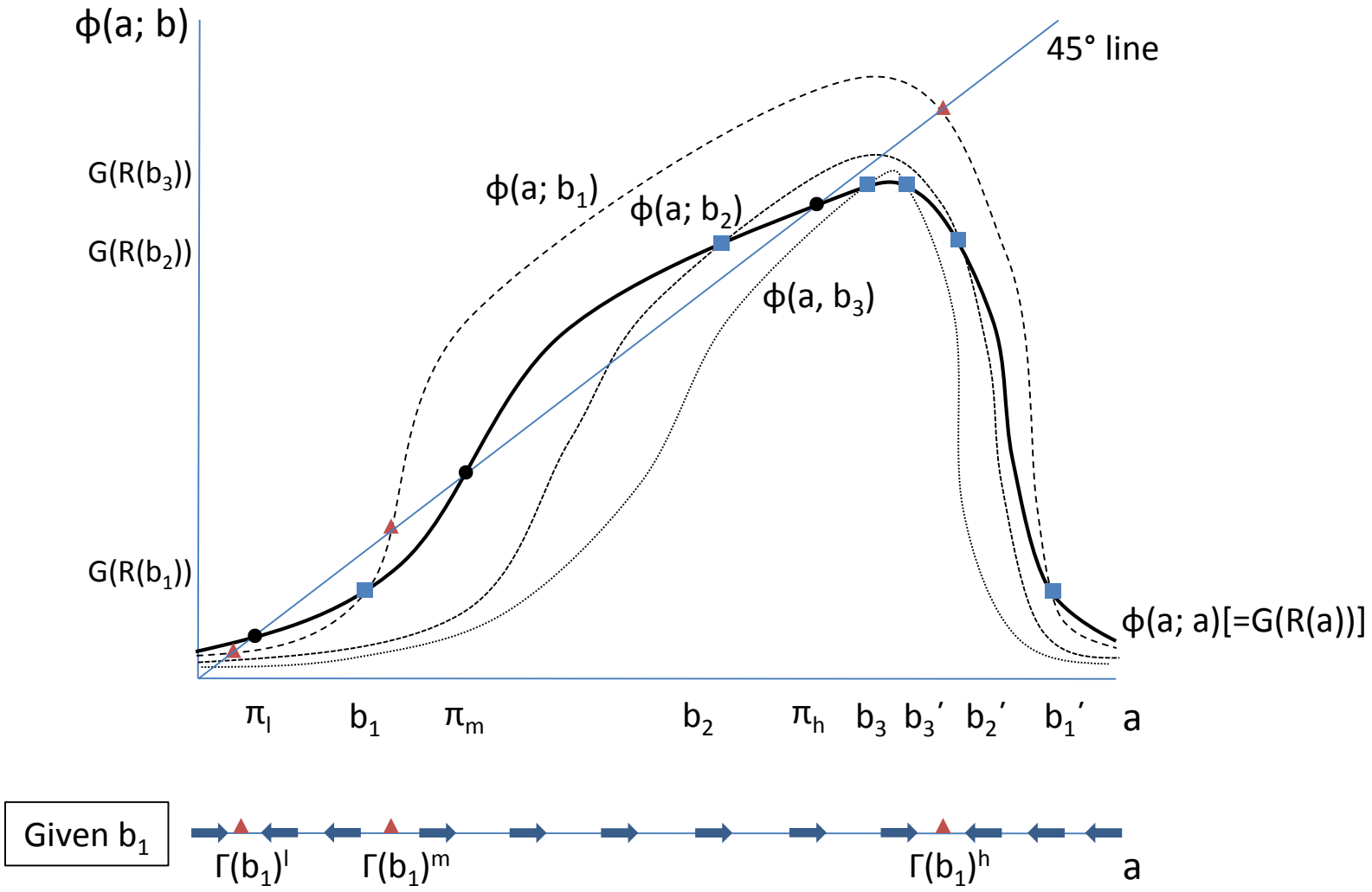
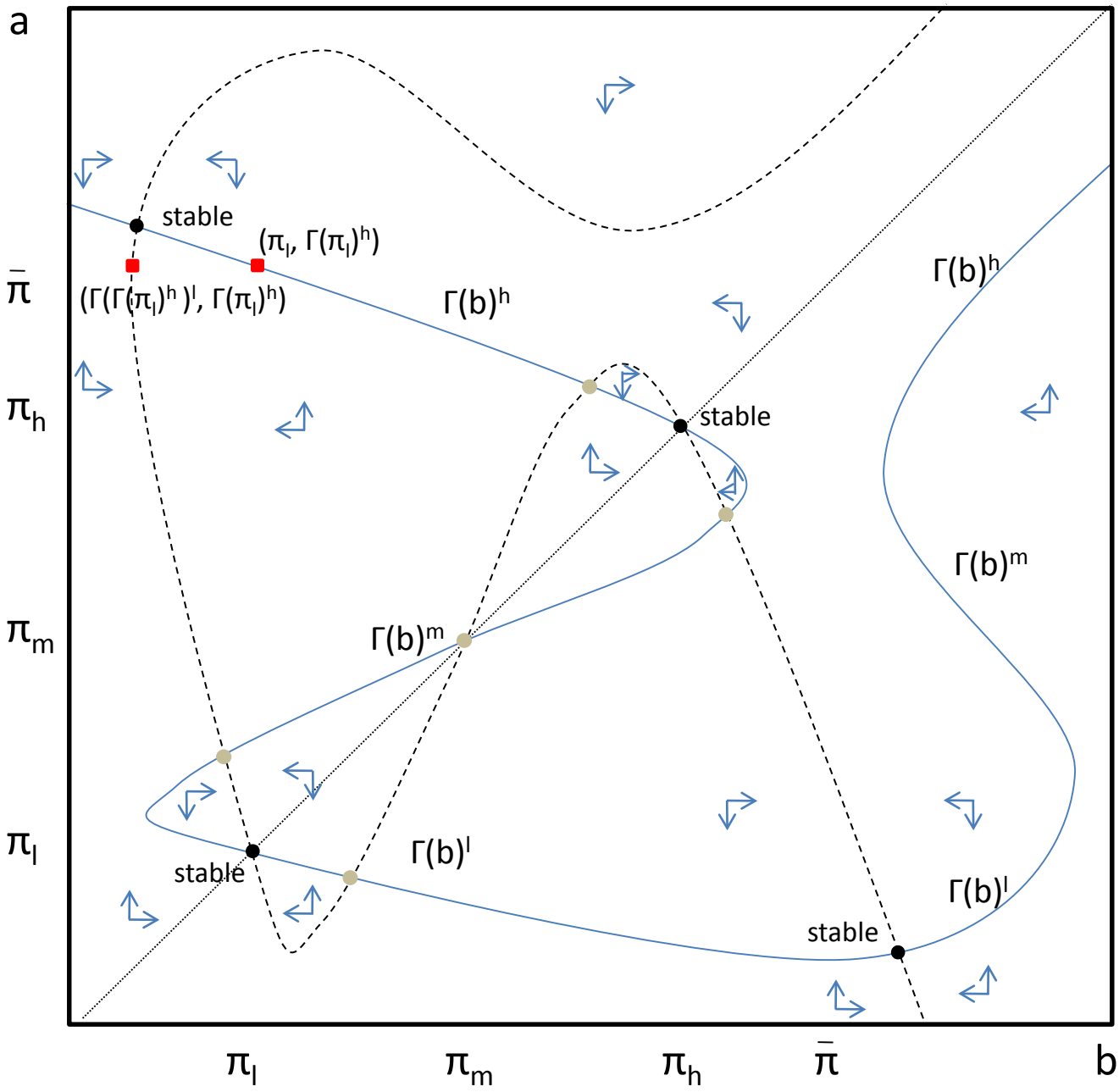
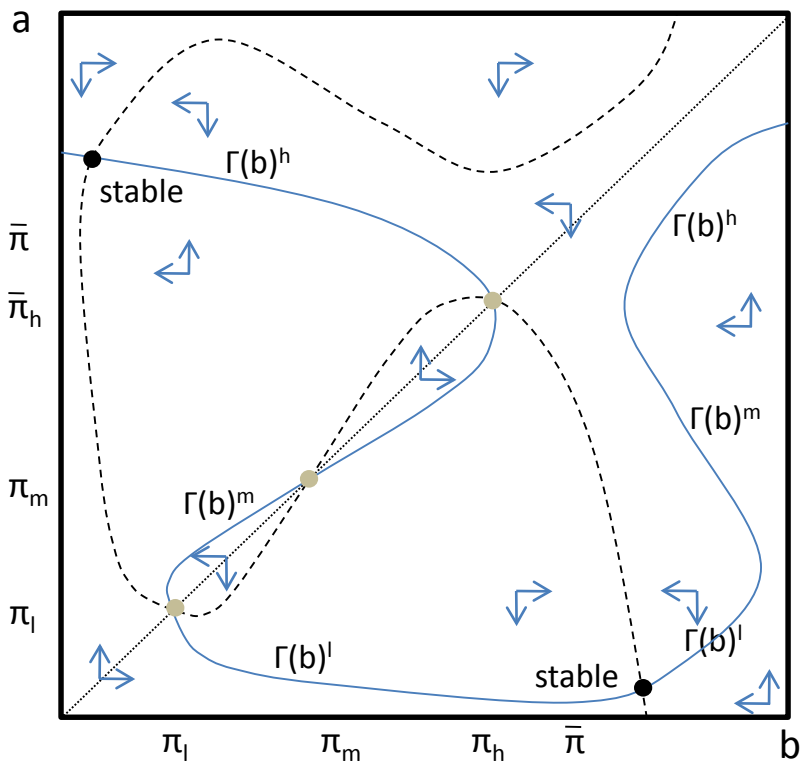


Figure 4. ESE given Multiple PSE (π_l, π_m, π_h)

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 $\Gamma'(\pi_l)>1$

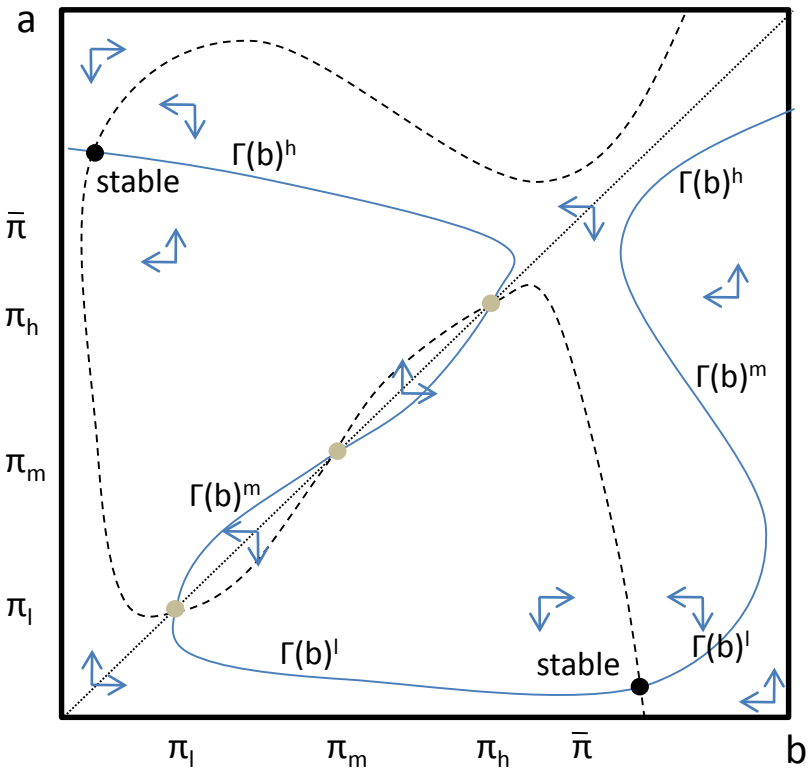
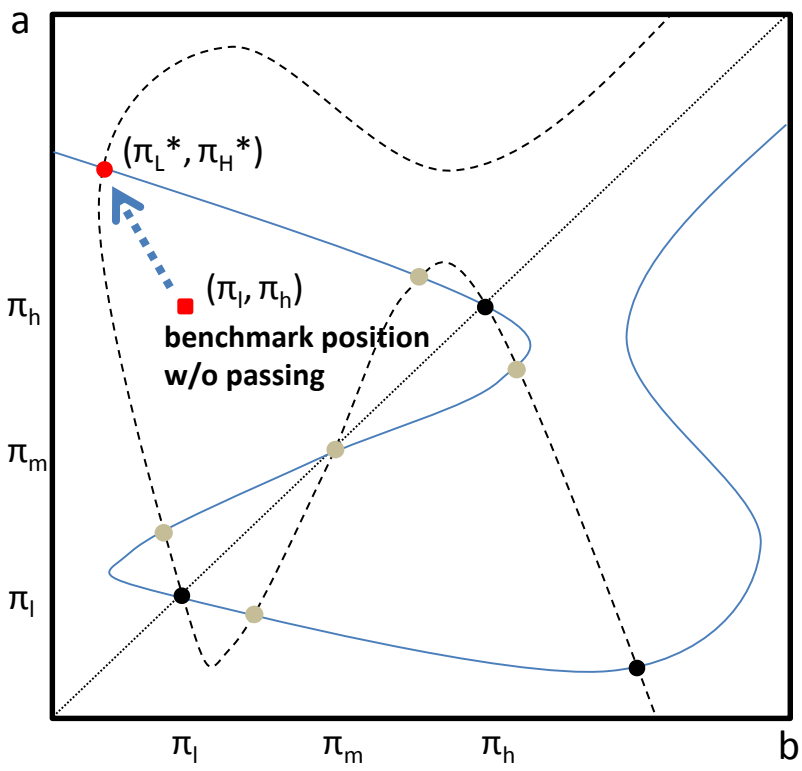
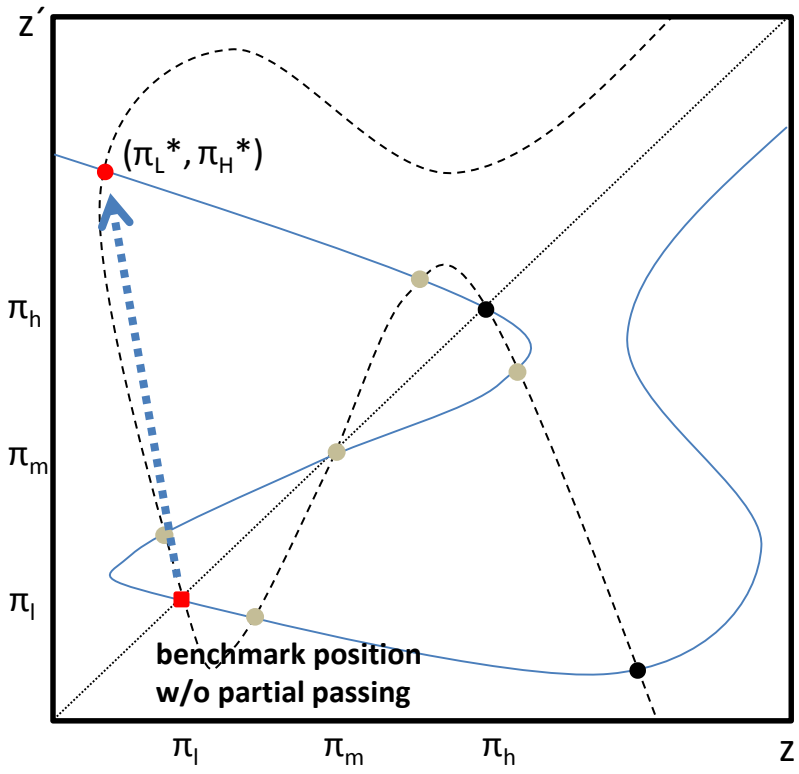


Figure 5. Passing and Partial Passing Behaviors

Panel A. Passing (Group B to Group A)



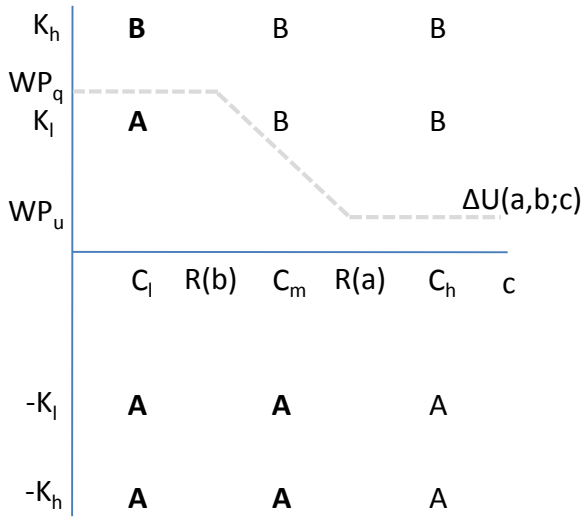
Panel B. Partial Passing (Subgroup Z to Subgroup Z´)



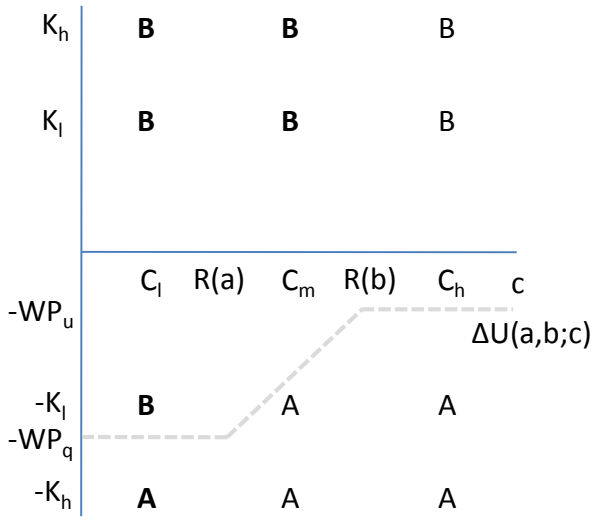
[For Online Publication]

Appendix Figure 1. PSE and ESE in a Simple Set-up

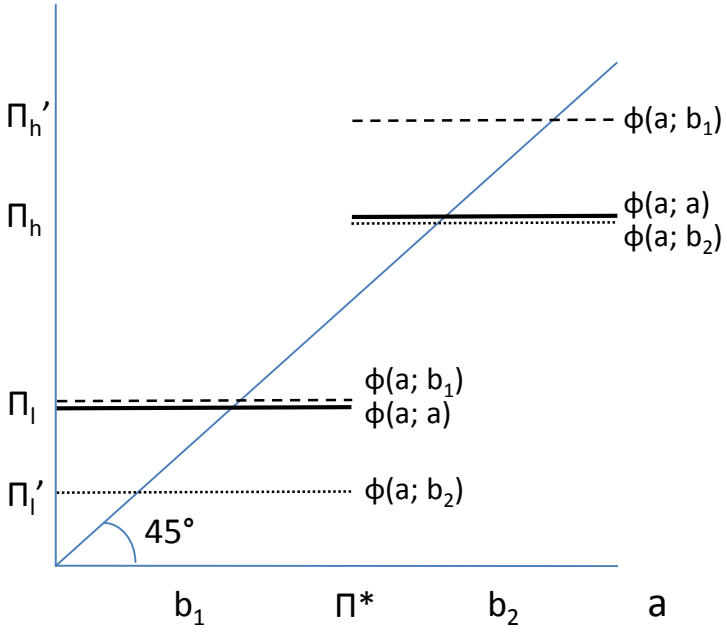
Panel A. ΔU(a,b;c) given b<Π*<a



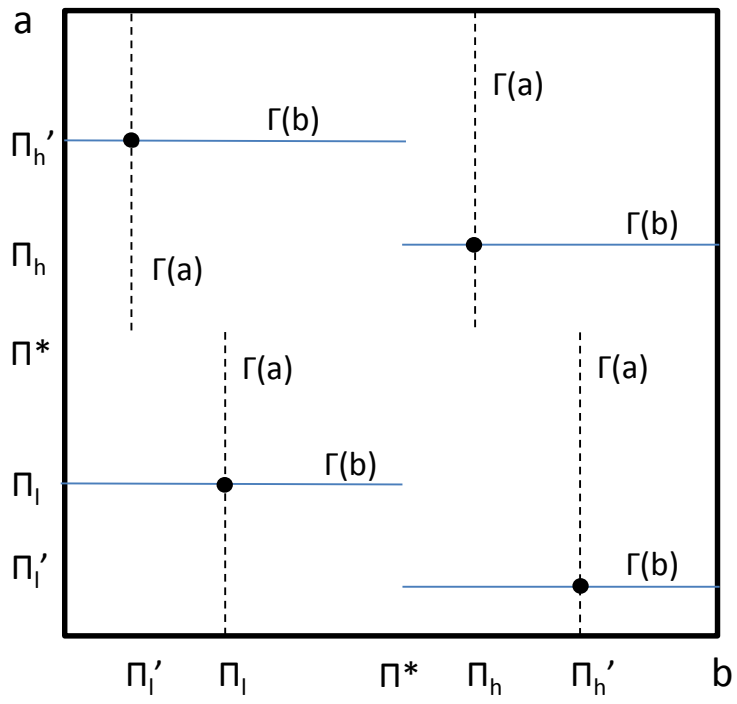
Panel B. ΔU(a,b;c) given a<Π*<b



Panel C. Investment Rate ϕ(a; b)



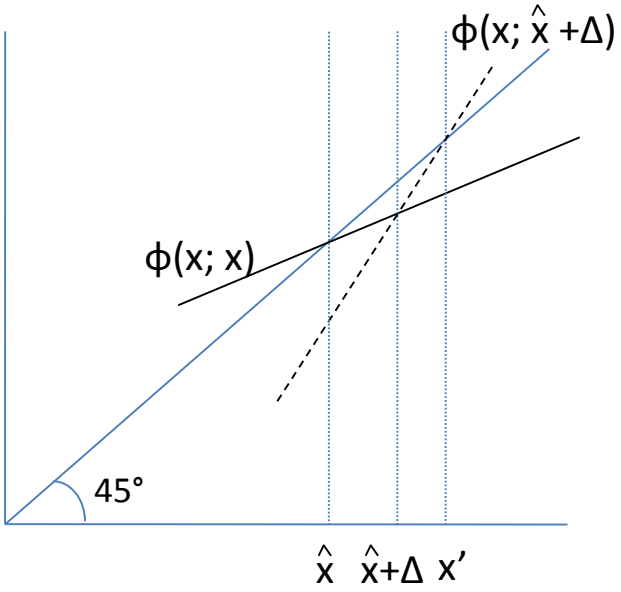
Panel D. ESE given Multiple PSE (Π_l, Π_h)



[For Online Publication]

Appendix Figure 2. Slope of Correspondence at Trivial ESE

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



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